DEX-0291

368 CD

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 August 2002 (22.08.2002)

PCT

(10) International Publication Number
WO 02/064611 A1

(51) International Patent Classification[7]: C07H 21/02, 21/04

(21) International Application Number: PCT/US02/04197

(22) International Filing Date: 12 February 2002 (12.02.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/268,292 13 February 2001 (13.02.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier application:
US 60/268,292 (CIP)
Filed on 13 February 2001 (13.02.2001)

(71) Applicant (*for all designated States except US*): DIADEXUS, INC. [US/US]; 343 Oyster Point Boulevard, South San Francisco, CA 94080 (US).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): SALCEDA, Susana [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). MACINA, Roberto, A. [AR/US]; 4118 Crescendo Avenue, San Jose, CA 95136 (US). HU, Ping [US/US]; 108 Palmer Street, San Ramon, CA 94583 (US). RECIPON, Herve [FR/US]; 85 Fortuna Avenue, San Francisco, CA 94115 (US). KARRA, Kalpana [IN/US]; 6511 Rainbow Drive, San Jose, CA 95129 (US). CAFFERKEY, Robert [IE/US]; 849 West Orange Avenue, Apartment 1030, South San Francisco, CA 94081 (US). SUN, Yongming [US/US]; 869 S. Winchester Boulevard, Apartment 260, San Jose, CA 95128 (US). LIU, Chenghua [CN/US]; 1125 Ranchero Way #14, San Jose, CA 95117 (US).

(74) Agents: LICATA, Jane, Massey et al.; Licata & Tyrrell P.C., 66 E. Main Street, Marlton, NJ 08053 (US).

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/064611 A1

(54) Title: COMPOSITIONS AND METHODS RELATING TO BREAST SPECIFIC GENES AND PROTEINS

(57) Abstract: The present invention relates to newly identified nucleic acids and polypeptides present in normal and neoplastic breast cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast tissue, identifying breast tissue, monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered breast tissue for treatment and research.

COMPOSITIONS AND METHODS RELATING TO BREAST SPECIFIC GENES AND PROTEINS

This application claims the benefit of priority from U.S. Provisional Application Serial No. 60/268,292 filed February 13, 2001, which is herein incorporated by reference in its entirety.

FIELD OF THE INVENTION

The present invention relates to newly identified nucleic acid molecules and polypeptides present in normal and neoplastic breast cells, including fragments, variants and derivatives of the nucleic acids and polypeptides. The present invention also relates to antibodies to the polypeptides of the invention, as well as agonists and antagonists of the polypeptides of the invention. The invention also relates to compositions comprising the nucleic acids, polypeptides, antibodies, variants, derivatives, agonists and antagonists of the invention and methods for the use of these compositions. These uses include identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast tissue, identifying breast tissue and monitoring and identifying and/or designing agonists and antagonists of polypeptides of the invention. The uses also include gene therapy, production of transgenic animals and cells, and production of engineered breast tissue for treatment and research.

BACKGROUND OF THE INVENTION

Excluding skin cancer, breast cancer, also called mammary tumor, is the most common cancer among women, accounting for a third of the cancers diagnosed in the United States. One in nine women will develop breast cancer in her lifetime and about 192,000 new cases of breast cancer are diagnosed annually with about 42,000 deaths. Bevers, *Primary Prevention of Breast Cancer*, in BREAST CANCER, 20-54 (Kelly K Hunt et al., ed., 2001); Kochanek et al., 49 Nat'l.Vital Statistics Reports 1, 14 (2001).

In the treatment of breast cancer, there is considerable emphasis on detection and risk assessment because early and accurate staging of breast cancer has a significant impact on survival. For example, breast cancer detected at an early stage (stage T0, discussed below) has a five-year survival rate of 92%. Conversely, if the cancer is not detected until a late stage (i.e., stage T4), the five-year survival rate is reduced to 13%. AJCC Cancer Staging Handbook pp. 164-65 (Irvin D. Fleming et al. eds., 5th ed. 1998). Some detection techniques, such as mammography and biopsy, involve increased

discomfort, expense, and/or radiation, and are only prescribed only to patients with an increased risk of breast cancer.

Current methods for predicting or detecting breast cancer risk are not optimal. One method for predicting the relative risk of breast cancer is by examining a patient's risk factors and pursuing aggressive diagnostic and treatment regiments for high risk patients. A patient's risk of breast cancer has been positively associated with increasing age, nulliparity, family history of breast cancer, personal history of breast cancer, early menarche, late menopause, late age of first full term pregnancy, prior proliferative breast disease, irradiation of the breast at an early age and a personal history of malignancy. Lifestyle factors such as fat consumption, alcohol consumption, education, and socioeconomic status have also been associated with an increased incidence of breast cancer although a direct cause and effect relationship has not been established. While these risk factors are statistically significant, their weak association with breast cancer limited their usefulness. Most women who develop breast cancer have none of the risk factors listed above, other than the risk that comes with growing older. NIH Publication No. 00-1556 (2000).

Current screening methods for detecting cancer, such as breast self exam, ultrasound, and mammography have drawbacks that reduce their effectiveness or prevent their widespread adoption. Breast self exams, while useful, are unreliable for the detection of breast cancer in the initial stages where the tumor is small and difficult to detect by palpitation. Ultrasound measurements require skilled operators at an increased expense. Mammography, while sensitive, is subject to over diagnosis in the detection of lesions that have questionable malignant potential. There is also the fear of the radiation used in mammography because prior chest radiation is a factor associated with an increase incidence of breast cancer.

At this time, there are no adequate methods of breast cancer prevention. The current methods of breast cancer prevention involve prophylactic mastectomy (mastectomy performed before cancer diagnosis) and chemoprevention (chemotherapy before cancer diagnosis) which are drastic measures that limit their adoption even among women with increased risk of breast cancer. Bevers, *supra*.

A number of genetic markers have been associated with breast cancer. Examples of these markers include carcinoembryonic antigen (CEA) (Mughal et al., 249 JAMA 1881 (1983)) MUC-1 (Frische and Liu, 22 J. Clin. Ligand 320 (2000)), HER-2/neu (Haris et al., 15 Proc.Am.Soc.Clin.Oncology. A96 (1996)), uPA, PAI-1, LPA, LPC,

RAK and BRCA (Esteva and Fritsche, *Serum and Tissue Markers for Breast Cancer*, in BREAST CANCER, 286-308 (2001)). These markers have problems with limited sensitivity, low correlation, and false negatives which limit their use for initial diagnosis. For example, while the BRCA1 gene mutation is useful as an indicator of an increased risk for breast cancer, it has limited use in cancer diagnosis because only 6.2 % of breast cancers are BRCA1 positive. Malone et al., 279 JAMA 922 (1998). *See also*, Mewman et al., 279 JAMA 915 (1998) (correlation of only 3.3%).

Breast cancers are diagnosed into the appropriate stage categories recognizing that different treatments are more effective for different stages of cancer. Stage TX indicates that primary tumor cannot be assessed (i.e., tumor was removed or breast tissue was removed). Stage T0 is characterized by abnormalities such as hyperplasia but with no evidence of primary tumor. Stage Tis is characterized by carcinoma in situ, intraductal carcinoma, lobular carcinoma in situ, or Paget's disease of the nipple with no tumor. Stage T1 is characterized as having a tumor of 2 cm or less in the greatest dimension. Within stage T1, Tmic indicates microinvasion of 0.1 cm or less, T1a indicates a tumor of between 0.1 to 0.5 cm, T1b indicates a tumor of between 0.5 to 1 cm, and T1c indicates tumors of between 1 cm to 2 cm. Stage T2 is characterized by tumors from 2 cm to 5 cm in the greatest dimension. Tumors greater than 5 cm in size are classified as stage T4. Within stage T4, T4a indicates extension of the tumor to the chess wall, T4b indicates edema or ulceration of the skin of the breast or satellite skin nodules confined to the same breast, T4c indicates a combination of T4a and T4b, and T4d indicates inflammatory carcinoma. <u>AJCC Cancer Staging Handbook</u> pp. 159-70 (Irvin D. Fleming et al. eds., 5th ed. 1998). In addition to standard staging, breast tumors may be classified according to their estrogen receptor and progesterone receptor protein status. Fisher et al., 7 Breast Cancer Research and Treatment 147 (1986). Additional pathological status, such as HER2/neu status may also be useful. Thor et al., 90 J.Nat'l.Cancer Inst. 1346 (1998); Paik et al., 90 J.Nat'l.Cancer Inst. 1361 (1998); Hutchins et al., 17 Proc.Am.Soc.Clin.Oncology A2 (1998).; and Simpson et al., 18 J.Clin.Oncology 2059 (2000).

In addition to the staging of the primary tumor, breast cancer metastases to regional lymph nodes may be staged. Stage NX indicates that the lymph nodes cannot be assessed (e.g., previously removed). Stage N0 indicates no regional lymph node metastasis. Stage N1 indicates metastasis to movable ipsilateral axillary lymph nodes. Stage N2 indicates metastasis to ipsilateral axillary lymph nodes fixed to one another or

to other structures. Stage N3 indicates metastasis to ipsilateral internal mammary lymph nodes. Id.

Stage determination has potential prognostic value and provides criteria for designing optimal therapy. Simpson et al., 18 J. Clin. Oncology 2059 (2000). Generally, pathological staging of breast cancer is preferable to clinical staging because the former gives a more accurate prognosis. However, clinical staging would be preferred if it were as accurate as pathological staging because it does not depend on an invasive procedure to obtain tissue for pathological evaluation. Staging of breast cancer would be improved by detecting new markers in cells, tissues, or bodily fluids which could differentiate between different stages of invasion. Progress in this field will allow more rapid and reliable method for treating breast cancer patients.

Treatment of breast cancer is generally decided after an accurate staging of the primary tumor. Primary treatment options include breast conserving therapy (lumpectomy, breast irradiation, and surgical staging of the axilla), and modified radical mastectomy. Additional treatments include chemotherapy, regional irradiation, and, in extreme cases, terminating estrogen production by ovarian ablation.

Until recently, the customary treatment for all breast cancer was mastectomy. Fonseca et al., 127 Annals of Internal Medicine 1013 (1997). However, recent data indicate that less radical procedures may be equally effective, in terms of survival, for early stage breast cancer. Fisher et al., 16 J. of Clinical Oncology 441 (1998). The treatment options for a patient with early stage breast cancer (i.e., stage Tis) may be breast-sparing surgery followed by localized radiation therapy at the breast. Alternatively, mastectomy optionally coupled with radiation or breast reconstruction may be employed. These treatment methods are equally effective in the early stages of breast cancer.

Patients with stage I and stage II breast cancer require surgery with chemotherapy and/or hormonal therapy. Surgery is of limited use in Stage III and stage IV patients. Thus, these patients are better candidates for chemotherapy and radiation therapy with surgery limited to biopsy to permit initial staging or subsequent restaging because cancer is rarely curative at this stage of the disease. AJCC Cancer Staging Handbook 84, ¶. 164-65 (Irvin D. Fleming et al. eds., 5th ed. 1998).

In an effort to provide more treatment options to patients, efforts are underway to define an earlier stage of breast cancer with low recurrence which could be treated with lumpectomy without postoperative radiation treatment. While a number of attempts have

been made to classify early stage breast cancer, no consensus recommendation on postoperative radiation treatment has been obtained from these studies. Page et al., 75 Cancer 1219 (1995); Fisher et al., 75 Cancer 1223 (1995); Silverstein et al., 77 Cancer 2267 (1996).

As discussed above, each of the methods for diagnosing and staging breast cancer is limited by the technology employed. Accordingly, there is need for sensitive molecular and cellular markers for the detection of breast cancer. There is a need for molecular markers for the accurate staging, including clinical and pathological staging, of breast cancers to optimize treatment methods. Finally, there is a need for sensitive molecular and cellular markers to monitor the progress of cancer treatments, including markers that can detect recurrence of breast cancers following remission.

Other objects, features, advantages and aspects of the present invention will become apparent to those of skill in the art from the following description. It should be understood, however, that the following description and the specific examples, while indicating preferred embodiments of the invention, are given by way of illustration only. Various changes and modifications within the spirit and scope of the disclosed invention will become readily apparent to those skilled in the art from reading the following description and from reading the other parts of the present disclosure.

SUMMARY OF THE INVENTION

The present invention solves these and other needs in the art by providing nucleic acid molecules and polypeptides as well as antibodies, agonists and antagonists, thereto that may be used to identify, diagnose, monitor, stage, image and treat breast cancer and non-cancerous disease states in breast; identify and monitor breast tissue; and identify and design agonists and antagonists of polypeptides of the invention. The invention also provides gene therapy, methods for producing transgenic animals and cells, and methods for producing engineered breast tissue for treatment and research.

Accordingly, one object of the invention is to provide nucleic acid molecules that are specific to breast cells and/or breast tissue. These breast specific nucleic acids (BSNAs) may be a naturally-occurring cDNA, genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. If the BSNA is genomic DNA, then the BSNA is a breast specific gene (BSG). In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to breast. In a more preferred embodiment, the nucleic acid molecule encodes a

polypeptide that comprises an amino acid sequence of SEQ ID NO: 172 through 295. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 171. By nucleic acid molecule, it is also meant to be inclusive of sequences that selectively hybridize or exhibit substantial sequence similarity to a nucleic acid molecule encoding a BSP, or that selectively hybridize or exhibit substantial sequence similarity to a BSNA, as well as allelic variants of a nucleic acid molecule encoding a BSP, and allelic variants of a BSNA. Nucleic acid molecules comprising a part of a nucleic acid sequence that encodes a BSP or that comprises a part of a nucleic acid sequence of a BSNA are also provided.

A related object of the present invention is to provide a nucleic acid molecule comprising one or more expression control sequences controlling the transcription and/or translation of all or a part of a BSNA. In a preferred embodiment, the nucleic acid molecule comprises one or more expression control sequences controlling the transcription and/or translation of a nucleic acid molecule that encodes all or a fragment of a BSP.

Another object of the invention is to provide vectors and/or host cells comprising a nucleic acid molecule of the instant invention. In a preferred embodiment, the nucleic acid molecule encodes all or a fragment of a BSP. In another preferred embodiment, the nucleic acid molecule comprises all or a part of a BSNA.

Another object of the invention is to provided methods for using the vectors and host cells comprising a nucleic acid molecule of the instant invention to recombinantly produce polypeptides of the invention.

Another object of the invention is to provide a polypeptide encoded by a nucleic acid molecule of the invention. In a preferred embodiment, the polypeptide is a BSP. The polypeptide may comprise either a fragment or a full-length protein as well as a mutant protein (mutein), fusion protein, homologous protein or a polypeptide encoded by an allelic variant of a BSP.

Another object of the invention is to provide an antibody that specifically binds to a polypeptide of the instant invention..

Another object of the invention is to provide agonists and antagonists of the nucleic acid molecules and polypeptides of the instant invention.

Another object of the invention is to provide methods for using the nucleic acid molecules to detect or amplify nucleic acid molecules that have similar or identical nucleic acid sequences compared to the nucleic acid molecules described herein. In a

preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying, diagnosing, monitoring, staging, imaging and treating breast cancer and non-cancerous disease states in breast. In another preferred embodiment, the invention provides methods of using the nucleic acid molecules of the invention for identifying and/or monitoring breast tissue. The nucleic acid molecules of the instant invention may also be used in gene therapy, for producing transgenic animals and cells, and for producing engineered breast tissue for treatment and research.

The polypeptides and/or antibodies of the instant invention may also be used to identify, diagnose, monitor, stage, image and treat breast cancer and non-cancerous disease states in breast. The invention provides methods of using the polypeptides of the invention to identify and/or monitor breast tissue, and to produce engineered breast tissue.

The agonists and antagonists of the instant invention may be used to treat breast cancer and non-cancerous disease states in breast and to produce engineered breast tissue.

Yet another object of the invention is to provide a computer readable means of storing the nucleic acid and amino acid sequences of the invention. The records of the computer readable means can be accessed for reading and displaying of sequences for comparison, alignment and ordering of the sequences of the invention to other sequences.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and General Techniques

Unless otherwise defined herein, scientific and technical terms used in connection with the present invention shall have the meanings that are commonly understood by those of ordinary skill in the art. Further, unless otherwise required by context, singular terms shall include pluralities and plural terms shall include the singular. Generally, nomenclatures used in connection with, and techniques of, cell and tissue culture, molecular biology, immunology, microbiology, genetics and protein and nucleic acid chemistry and hybridization described herein are those well-known and commonly used in the art. The methods and techniques of the present invention are generally performed according to conventional methods well-known in the art and as described in various general and more specific references that are cited and discussed throughout the present specification unless otherwise indicated. *See, e.g.*, Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 2d ed., Cold Spring Harbor Laboratory Press (1989) and Sambrook *et al.*, Molecular Cloning: A Laboratory Manual, 3d ed., Cold Spring Harbor

Press (2001); Ausubel *et al.*, Current Protocols in Molecular Biology, Greene Publishing Associates (1992, and Supplements to 2000); Ausubel *et al.*, Short Protocols in Molecular Biology: A Compendium of Methods from Current Protocols in Molecular Biology – 4th Ed., Wiley & Sons (1999); Harlow and Lane, Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1990); and Harlow and Lane, Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory Press (1999); each of which is incorporated herein by reference in its entirety.

Enzymatic reactions and purification techniques are performed according to manufacturer's specifications, as commonly accomplished in the art or as described herein. The nomenclatures used in connection with, and the laboratory procedures and techniques of, analytical chemistry, synthetic organic chemistry, and medicinal and pharmaceutical chemistry described herein are those well-known and commonly used in the art. Standard techniques are used for chemical syntheses, chemical analyses, pharmaceutical preparation, formulation, and delivery, and treatment of patients.

The following terms, unless otherwise indicated, shall be understood to have the following meanings:

A "nucleic acid molecule" of this invention refers to a polymeric form of nucleotides and includes both sense and antisense strands of RNA, cDNA, genomic DNA, and synthetic forms and mixed polymers of the above. A nucleotide refers to a ribonucleotide, deoxynucleotide or a modified form of either type of nucleotide. A "nucleic acid molecule" as used herein is synonymous with "nucleic acid" and "polynucleotide." The term "nucleic acid molecule" usually refers to a molecule of at least 10 bases in length, unless otherwise specified. The term includes single- and double-stranded forms of DNA. In addition, a polynucleotide may include either or both naturally-occurring and modified nucleotides linked together by naturally-occurring and/or non-naturally occurring nucleotide linkages.

The nucleic acid molecules may be modified chemically or biochemically or may contain non-natural or derivatized nucleotide bases, as will be readily appreciated by those of skill in the art. Such modifications include, for example, labels, methylation, substitution of one or more of the naturally occurring nucleotides with an analog, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoramidates, carbamates, etc.), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates, etc.), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen, etc.), chelators, alkylators, and modified linkages

(*e.g.*, alpha anomeric nucleic acids, etc.) The term "nucleic acid molecule" also includes any topological conformation, including single-stranded, double-stranded, partially duplexed, triplexed, hairpinned, circular and padlocked conformations. Also included are synthetic molecules that mimic polynucleotides in their ability to bind to a designated sequence via hydrogen bonding and other chemical interactions. Such molecules are known in the art and include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

A "gene" is defined as a nucleic acid molecule that comprises a nucleic acid sequence that encodes a polypeptide and the expression control sequences that surround the nucleic acid sequence that encodes the polypeptide. For instance, a gene may comprise a promoter, one or more enhancers, a nucleic acid sequence that encodes a polypeptide, downstream regulatory sequences and, possibly, other nucleic acid sequences involved in regulation of the expression of an RNA. As is well-known in the art, eukaryotic genes usually contain both exons and introns. The term "exon" refers to a nucleic acid sequence found in genomic DNA that is bioinformatically predicted and/or experimentally confirmed to contribute a contiguous sequence to a mature mRNA transcript. The term "intron" refers to a nucleic acid sequence found in genomic DNA that is predicted and/or confirmed to not contribute to a mature mRNA transcript, but rather to be "spliced out" during processing of the transcript.

A nucleic acid molecule or polypeptide is "derived" from a particular species if the nucleic acid molecule or polypeptide has been isolated from the particular species, or if the nucleic acid molecule or polypeptide is homologous to a nucleic acid molecule or polypeptide isolated from a particular species.

An "isolated" or "substantially pure" nucleic acid or polynucleotide (*e.g.*, an RNA, DNA or a mixed polymer) is one which is substantially separated from other cellular components that naturally accompany the native polynucleotide in its natural host cell, *e.g.*, ribosomes, polymerases, or genomic sequences with which it is naturally associated. The term embraces a nucleic acid or polynucleotide that (1) has been removed from its naturally occurring environment, (2) is not associated with all or a portion of a polynucleotide in which the "isolated polynucleotide" is found in nature, (3) is operatively linked to a polynucleotide which it is not linked to in nature, (4) does not occur in nature as part of a larger sequence or (5) includes nucleotides or internucleoside bonds that are not found in nature. The term "isolated" or "substantially pure" also can be used in reference to recombinant or cloned DNA isolates, chemically synthesized

polynucleotide analogs, or polynucleotide analogs that are biologically synthesized by heterologous systems. The term "isolated nucleic acid molecule" includes nucleic acid molecules that are integrated into a host cell chromosome at a heterologous site, recombinant fusions of a native fragment to a heterologous sequence, recombinant vectors present as episomes or as integrated into a host cell chromosome.

A "part" of a nucleic acid molecule refers to a nucleic acid molecule that comprises a partial contiguous sequence of at least 10 bases of the reference nucleic acid molecule. Preferably, a part comprises at least 15 to 20 bases of a reference nucleic acid molecule. In theory, a nucleic acid sequence of 17 nucleotides is of sufficient length to occur at random less frequently than once in the three gigabase human genome, and thus to provide a nucleic acid probe that can uniquely identify the reference sequence in a nucleic acid mixture of genomic complexity. A preferred part is one that comprises a nucleic acid sequence that can encode at least 6 contiguous amino acid sequences (fragments of at least 18 nucleotides) because they are useful in directing the expression or synthesis of peptides that are useful in mapping the epitopes of the polypeptide encoded by the reference nucleic acid. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81:3998-4002 (1984); and United States Patent Nos. 4,708,871 and 5,595,915, the disclosures of which are incorporated herein by reference in their entireties. A part may also comprise at least 25, 30, 35 or 40 nucleotides of a reference nucleic acid molecule, or at least 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides of a reference nucleic acid molecule. A part of a nucleic acid molecule may comprise no other nucleic acid sequences. Alternatively, a part of a nucleic acid may comprise other nucleic acid sequences from other nucleic acid molecules.

The term "oligonucleotide" refers to a nucleic acid molecule generally comprising a length of 200 bases or fewer. The term often refers to single-stranded deoxyribonucleotides, but it can refer as well to single- or double-stranded ribonucleotides, RNA:DNA hybrids and double-stranded DNAs, among others. Preferably, oligonucleotides are 10 to 60 bases in length and most preferably 12, 13, 14, 15, 16, 17, 18, 19 or 20 bases in length. Other preferred oligonucleotides are 25, 30, 35, 40, 45, 50, 55 or 60 bases in length. Oligonucleotides may be single-stranded, *e.g.* for use as probes or primers, or may be double-stranded, *e.g.* for use in the construction of a mutant gene. Oligonucleotides of the invention can be either sense or antisense oligonucleotides. An oligonucleotide can be derivatized or modified as discussed above for nucleic acid molecules.

Oligonucleotides, such as single-stranded DNA probe oligonucleotides, often are synthesized by chemical methods, such as those implemented on automated oligonucleotide synthesizers. However, oligonucleotides can be made by a variety of other methods, including *in vitro* recombinant DNA-mediated techniques and by expression of DNAs in cells and organisms. Initially, chemically synthesized DNAs typically are obtained without a 5' phosphate. The 5' ends of such oligonucleotides are not substrates for phosphodiester bond formation by ligation reactions that employ DNA ligases typically used to form recombinant DNA molecules. Where ligation of such oligonucleotides is desired, a phosphate can be added by standard techniques, such as those that employ a kinase and ATP. The 3' end of a chemically synthesized oligonucleotide generally has a free hydroxyl group and, in the presence of a ligase, such as T4 DNA ligase, readily will form a phosphodiester bond with a 5' phosphate of another polynucleotide, such as another oligonucleotide. As is well-known, this reaction can be prevented selectively, where desired, by removing the 5' phosphates of the other polynucleotide(s) prior to ligation.

The term "naturally-occurring nucleotide" referred to herein includes naturally-occurring deoxyribonucleotides and ribonucleotides. The term "modified nucleotides" referred to herein includes nucleotides with modified or substituted sugar groups and the like. The term "nucleotide linkages" referred to herein includes nucleotides linkages such as phosphorothioate, phosphorodithioate, phosphoroselenoate, phosphorodiselenoate, phosphoroanilothioate, phoshoraniladate, phosphoroamidate, and the like. *See e.g.*, LaPlanche *et al. Nucl. Acids Res.* 14:9081-9093 (1986); Stein *et al. Nucl. Acids Res.* 16:3209-3221 (1988); Zon *et al. Anti-Cancer Drug Design* 6:539-568 (1991); Zon *et al.*, in Eckstein (ed.) Oligonucleotides and Analogues: A Practical Approach, pp. 87-108, Oxford University Press (1991); United States Patent No. 5,151,510; Uhlmann and Peyman *Chemical Reviews* 90:543 (1990), the disclosures of which are hereby incorporated by reference.

Unless specified otherwise, the left hand end of a polynucleotide sequence in sense orientation is the 5' end and the right hand end of the sequence is the 3' end. In addition, the left hand direction of a polynucleotide sequence in sense orientation is referred to as the 5' direction, while the right hand direction of the polynucleotide sequence is referred to as the 3' direction. Further, unless otherwise indicated, each nucleotide sequence is set forth herein as a sequence of deoxyribonucleotides. It is intended, however, that the given sequence be interpreted as would be appropriate to the

polynucleotide composition: for example, if the isolated nucleic acid is composed of RNA, the given sequence intends ribonucleotides, with uridine substituted for thymidine.

The term "allelic variant" refers to one of two or more alternative naturally-occurring forms of a gene, wherein each gene possesses a unique nucleotide sequence. In a preferred embodiment, different alleles of a given gene have similar or identical biological properties.

The term "percent sequence identity" in the context of nucleic acid sequences refers to the residues in two sequences which are the same when aligned for maximum correspondence. The length of sequence identity comparison may be over a stretch of at least about nine nucleotides, usually at least about 20 nucleotides, more usually at least about 24 nucleotides, typically at least about 28 nucleotides, more typically at least about 32 nucleotides, and preferably at least about 36 or more nucleotides. There are a number of different algorithms known in the art which can be used to measure nucleotide sequence identity. For instance, polynucleotide sequences can be compared using FASTA, Gap or Bestfit, which are programs in Wisconsin Package Version 10.0, Genetics Computer Group (GCG), Madison, Wisconsin. FASTA, which includes, *e.g.*, the programs FASTA2 and FASTA3, provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson, *Methods Enzymol.* 183: 63-98 (1990); Pearson, *Methods Mol. Biol.* 132: 185-219 (2000); Pearson, *Methods Enzymol.* 266: 227-258 (1996); Pearson, *J. Mol. Biol.* 276: 71-84 (1998); herein incorporated by reference). Unless otherwise specified, default parameters for a particular program or algorithm are used. For instance, percent sequence identity between nucleic acid sequences can be determined using FASTA with its default parameters (a word size of 6 and the NOPAM factor for the scoring matrix) or using Gap with its default parameters as provided in GCG Version 6.1, herein incorporated by reference.

A reference to a nucleic acid sequence encompasses its complement unless otherwise specified. Thus, a reference to a nucleic acid molecule having a particular sequence should be understood to encompass its complementary strand, with its complementary sequence. The complementary strand is also useful, *e.g.*, for antisense therapy, hybridization probes and PCR primers.

In the molecular biology art, researchers use the terms "percent sequence identity", "percent sequence similarity" and "percent sequence homology"

interchangeably. In this application, these terms shall have the same meaning with respect to nucleic acid sequences only.

The term "substantial similarity" or "substantial sequence similarity," when referring to a nucleic acid or fragment thereof, indicates that, when optimally aligned with appropriate nucleotide insertions or deletions with another nucleic acid (or its complementary strand), there is nucleotide sequence identity in at least about 50%, more preferably 60% of the nucleotide bases, usually at least about 70%, more usually at least about 80%, preferably at least about 90%, and more preferably at least about 95-98% of the nucleotide bases, as measured by any well-known algorithm of sequence identity, such as FASTA, BLAST or Gap, as discussed above.

Alternatively, substantial similarity exists when a nucleic acid or fragment thereof hybridizes to another nucleic acid, to a strand of another nucleic acid, or to the complementary strand thereof, under selective hybridization conditions. Typically, selective hybridization will occur when there is at least about 55% sequence identity, preferably at least about 65%, more preferably at least about 75%, and most preferably at least about 90% sequence identity, over a stretch of at least about 14 nucleotides, more preferably at least 17 nucleotides, even more preferably at least 20, 25, 30, 35, 40, 50, 60, 70, 80, 90 or 100 nucleotides.

Nucleic acid hybridization will be affected by such conditions as salt concentration, temperature, solvents, the base composition of the hybridizing species, length of the complementary regions, and the number of nucleotide base mismatches between the hybridizing nucleic acids, as will be readily appreciated by those skilled in the art. "Stringent hybridization conditions" and "stringent wash conditions" in the context of nucleic acid hybridization experiments depend upon a number of different physical parameters. The most important parameters include temperature of hybridization, base composition of the nucleic acids, salt concentration and length of the nucleic acid. One having ordinary skill in the art knows how to vary these parameters to achieve a particular stringency of hybridization. In general, "stringent hybridization" is performed at about 25°C below the thermal melting point (T_m) for the specific DNA hybrid under a particular set of conditions. "Stringent washing" is performed at temperatures about 5°C lower than the T_m for the specific DNA hybrid under a particular set of conditions. The T_m is the temperature at which 50% of the target sequence hybridizes to a perfectly matched probe. *See* Sambrook (1989), *supra*, p. 9.51, hereby incorporated by reference.

The T_m for a particular DNA-DNA hybrid can be estimated by the formula:
$T_m = 81.5°C + 16.6 (\log_{10}[Na^+]) + 0.41$ (fraction G + C) – 0.63 (% formamide) - (600/l) where l is the length of the hybrid in base pairs.

The T_m for a particular RNA-RNA hybrid can be estimated by the formula:
$T_m = 79.8°C + 18.5 (\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.35 (% formamide) - (820/l).

The T_m for a particular RNA-DNA hybrid can be estimated by the formula:
$T_m = 79.8°C + 18.5(\log_{10}[Na^+]) + 0.58$ (fraction G + C) + 11.8 (fraction G + C)^2 - 0.50 (% formamide) - (820/l).

In general, the T_m decreases by 1-1.5°C for each 1% of mismatch between two nucleic acid sequences. Thus, one having ordinary skill in the art can alter hybridization and/or washing conditions to obtain sequences that have higher or lower degrees of sequence identity to the target nucleic acid. For instance, to obtain hybridizing nucleic acids that contain up to 10% mismatch from the target nucleic acid sequence, 10-15°C would be subtracted from the calculated T_m of a perfectly matched hybrid, and then the hybridization and washing temperatures adjusted accordingly. Probe sequences may also hybridize specifically to duplex DNA under certain conditions to form triplex or other higher order DNA complexes. The preparation of such probes and suitable hybridization conditions are well-known in the art.

An example of stringent hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 50% formamide/6X SSC at 42°C for at least ten hours and preferably overnight (approximately 16 hours). Another example of stringent hybridization conditions is 6X SSC at 68°C without formamide for at least ten hours and preferably overnight. An example of moderate stringency hybridization conditions is 6X SSC at 55°C without formamide for at least ten hours and preferably overnight. An example of low stringency hybridization conditions for hybridization of complementary nucleic acid sequences having more than 100 complementary residues on a filter in a Southern or Northern blot or for screening a library is 6X SSC at 42°C for at least ten hours. Hybridization conditions to identify nucleic acid sequences that are similar but not identical can be identified by experimentally changing the hybridization temperature from 68°C to 42°C while keeping the salt concentration constant (6X SSC), or keeping the hybridization temperature and salt concentration constant (*e.g.* 42°C and 6X SSC) and varying the formamide

concentration from 50% to 0%. Hybridization buffers may also include blocking agents to lower background. These agents are well-known in the art. *See* Sambrook *et al.* (1989), *supra*, pages 8.46 and 9.46-9.58, herein incorporated by reference. *See also* Ausubel (1992), *supra*, Ausubel (1999), *supra*, and Sambrook (2001), *supra*.

Wash conditions also can be altered to change stringency conditions. An example of stringent wash conditions is a 0.2x SSC wash at 65°C for 15 minutes (*see* Sambrook (1989), *supra*, for SSC buffer). Often the high stringency wash is preceded by a low stringency wash to remove excess probe. An exemplary medium stringency wash for duplex DNA of more than 100 base pairs is 1x SSC at 45°C for 15 minutes. An exemplary low stringency wash for such a duplex is 4x SSC at 40°C for 15 minutes. In general, signal-to-noise ratio of 2x or higher than that observed for an unrelated probe in the particular hybridization assay indicates detection of a specific hybridization.

As defined herein, nucleic acid molecules that do not hybridize to each other under stringent conditions are still substantially similar to one another if they encode polypeptides that are substantially identical to each other. This occurs, for example, when a nucleic acid molecule is created synthetically or recombinantly using high codon degeneracy as permitted by the redundancy of the genetic code.

Hybridization conditions for nucleic acid molecules that are shorter than 100 nucleotides in length (*e.g.*, for oligonucleotide probes) may be calculated by the formula:

$T_m = 81.5°C + 16.6(\log_{10}[Na^+]) + 0.41(\text{fraction G+C}) -(600/N)$,

wherein N is change length and the $[Na^+]$ is 1 M or less. *See* Sambrook (1989), *supra*, p. 11.46. For hybridization of probes shorter than 100 nucleotides, hybridization is usually performed under stringent conditions (5-10°C below the T_m) using high concentrations (0.1-1.0 pmol/ml) of probe. *Id.* at p. 11.45. Determination of hybridization using mismatched probes, pools of degenerate probes or "guessmers," as well as hybridization solutions and methods for empirically determining hybridization conditions are well-known in the art. *See, e.g.*, Ausubel (1999), *supra*; Sambrook (1989), *supra*, pp. 11.45-11.57.

The term "digestion" or "digestion of DNA" refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes referred to herein are commercially available and their reaction conditions, cofactors and other requirements for use are known and routine to the skilled artisan. For analytical purposes, typically, 1 µg of plasmid or DNA fragment is digested with about 2 units of enzyme in about 20 µl of reaction buffer. For the

purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 μg of DNA are digested with 20 to 250 units of enzyme in proportionately larger volumes. Appropriate buffers and substrate amounts for particular restriction enzymes are described in standard laboratory manuals, such as those referenced below, and they are specified by commercial suppliers. Incubation times of about 1 hour at 37°C are ordinarily used, but conditions may vary in accordance with standard procedures, the supplier's instructions and the particulars of the reaction. After digestion, reactions may be analyzed, and fragments may be purified by electrophoresis through an agarose or polyacrylamide gel, using well-known methods that are routine for those skilled in the art.

The term "ligation" refers to the process of forming phosphodiester bonds between two or more polynucleotides, which most often are double-stranded DNAS. Techniques for ligation are well-known to the art and protocols for ligation are described in standard laboratory manuals and references, such as, *e.g.*, Sambrook (1989), *supra*.

Genome-derived "single exon probes," are probes that comprise at least part of an exon ("reference exon") and can hybridize detectably under high stringency conditions to transcript-derived nucleic acids that include the reference exon but do not hybridize detectably under high stringency conditions to nucleic acids that lack the reference exon. Single exon probes typically further comprise, contiguous to a first end of the exon portion, a first intronic and/or intergenic sequence that is identically contiguous to the exon in the genome, and may contain a second intronic and/or intergenic sequence that is identically contiguous to the exon in the genome. The minimum length of genome-derived single exon probes is defined by the requirement that the exonic portion be of sufficient length to hybridize under high stringency conditions to transcript-derived nucleic acids, as discussed above. The maximum length of genome-derived single exon probes is defined by the requirement that the probes contain portions of no more than one exon. The single exon probes may contain priming sequences not found in contiguity with the rest of the probe sequence in the genome, which priming sequences are useful for PCR and other amplification-based technologies.

The term "microarray" or "nucleic acid microarray" refers to a substrate-bound collection of plural nucleic acids, hybridization to each of the plurality of bound nucleic acids being separately detectable. The substrate can be solid or porous, planar or non-planar, unitary or distributed. Microarrays or nucleic acid microarrays include all the devices so called in Schena (ed.), DNA Microarrays: A Practical Approach (Practical

Approach Series), Oxford University Press (1999); *Nature Genet.* 21(1)(suppl.):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000). These microarrays include substrate-bound collections of plural nucleic acids in which the plurality of nucleic acids are disposed on a plurality of beads, rather than on a unitary planar substrate, as is described, *inter alia*, in Brenner *et al.*, *Proc. Natl. Acad. Sci. USA* 97(4):1665-1670 (2000).

The term "mutated" when applied to nucleic acid molecules means that nucleotides in the nucleic acid sequence of the nucleic acid molecule may be inserted, deleted or changed compared to a reference nucleic acid sequence. A single alteration may be made at a locus (a point mutation) or multiple nucleotides may be inserted, deleted or changed at a single locus. In addition, one or more alterations may be made at any number of loci within a nucleic acid sequence. In a preferred embodiment, the nucleic acid molecule comprises the wild type nucleic acid sequence encoding a BSP or is a BSNA. The nucleic acid molecule may be mutated by any method known in the art including those mutagenesis techniques described *infra*.

The term "error-prone PCR" refers to a process for performing PCR under conditions where the copying fidelity of the DNA polymerase is low, such that a high rate of point mutations is obtained along the entire length of the PCR product. *See, e.g.*, Leung *et al.*, *Technique* 1: 11-15 (1989) and Caldwell *et al.*, *PCR Methods Applic.* 2: 28-33 (1992).

The term "oligonucleotide-directed mutagenesis" refers to a process which enables the generation of site-specific mutations in any cloned DNA segment of interest. *See, e.g.*, Reidhaar-Olson *et al.*, *Science* 241: 53-57 (1988).

The term "assembly PCR" refers to a process which involves the assembly of a PCR product from a mixture of small DNA fragments. A large number of different PCR reactions occur in parallel in the same vial, with the products of one reaction priming the products of another reaction.

The term "sexual PCR mutagenesis" or "DNA shuffling" refers to a method of error-prone PCR coupled with forced homologous recombination between DNA molecules of different but highly related DNA sequence *in vitro*, caused by random fragmentation of the DNA molecule based on sequence similarity, followed by fixation of the crossover by primer extension in an error-prone PCR reaction. *See, e.g.*, Stemmer, *Proc. Natl. Acad. Sci. U.S.A.* 91: 10747-10751 (1994). DNA shuffling can be carried out between several related genes ("Family shuffling").

The term "*in vivo* mutagenesis" refers to a process of generating random mutations in any cloned DNA of interest which involves the propagation of the DNA in a strain of bacteria such as *E. coli* that carries mutations in one or more of the DNA repair pathways. These "mutator" strains have a higher random mutation rate than that of a wild-type parent. Propagating the DNA in a mutator strain will eventually generate random mutations within the DNA.

The term "cassette mutagenesis" refers to any process for replacing a small region of a double-stranded DNA molecule with a synthetic oligonucleotide "cassette" that differs from the native sequence. The oligonucleotide often contains completely and/or partially randomized native sequence.

The term "recursive ensemble mutagenesis" refers to an algorithm for protein engineering (protein mutagenesis) developed to produce diverse populations of phenotypically related mutants whose members differ in amino acid sequence. This method uses a feedback mechanism to control successive rounds of combinatorial cassette mutagenesis. *See, e.g.*, Arkin *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* 89: 7811-7815 (1992).

The term "exponential ensemble mutagenesis" refers to a process for generating combinatorial libraries with a high percentage of unique and functional mutants, wherein small groups of residues are randomized in parallel to identify, at each altered position, amino acids which lead to functional proteins. *See, e.g.*, Delegrave *et al.*, *Biotechnology Research* 11: 1548-1552 (1993); Arnold, *Current Opinion in Biotechnology* 4: 450-455 (1993). Each of the references mentioned above are hereby incorporated by reference in its entirety.

"Operatively linked" expression control sequences refers to a linkage in which the expression control sequence is contiguous with the gene of interest to control the gene of interest, as well as expression control sequences that act in *trans* or at a distance to control the gene of interest.

The term "expression control sequence" as used herein refers to polynucleotide sequences which are necessary to affect the expression of coding sequences to which they are operatively linked. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Expression control sequences include appropriate transcription initiation, termination, promoter and enhancer sequences; efficient RNA processing signals such as splicing and polyadenylation signals; sequences that stabilize cytoplasmic mRNA; sequences that

enhance translation efficiency (*e.g.*, ribosome binding sites); sequences that enhance protein stability; and when desired, sequences that enhance protein secretion. The nature of such control sequences differs depending upon the host organism; in prokaryotes, such control sequences generally include the promoter, ribosomal binding site, and transcription termination sequence. The term "control sequences" is intended to include, at a minimum, all components whose presence is essential for expression, and can also include additional components whose presence is advantageous, for example, leader sequences and fusion partner sequences.

The term "vector," as used herein, is intended to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double-stranded DNA loop into which additional DNA segments may be ligated. Other vectors include cosmids, bacterial artificial chromosomes (BAC) and yeast artificial chromosomes (YAC). Another type of vector is a viral vector, wherein additional DNA segments may be ligated into the viral genome. Viral vectors that infect bacterial cells are referred to as bacteriophages. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication). Other vectors can be integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "recombinant expression vectors" (or simply, "expression vectors"). In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and "vector" may be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include other forms of expression vectors that serve equivalent functions.

The term "recombinant host cell" (or simply "host cell"), as used herein, is intended to refer to a cell into which an expression vector has been introduced. It should be understood that such terms are intended to refer not only to the particular subject cell but to the progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term "host cell" as used herein.

As used herein, the phrase "open reading frame" and the equivalent acronym "ORF" refer to that portion of a transcript-derived nucleic acid that can be translated in its entirety into a sequence of contiguous amino acids. As so defined, an ORF has length, measured in nucleotides, exactly divisible by 3. As so defined, an ORF need not encode the entirety of a natural protein.

As used herein, the phrase "ORF-encoded peptide" refers to the predicted or actual translation of an ORF.

As used herein, the phrase "degenerate variant" of a reference nucleic acid sequence intends all nucleic acid sequences that can be directly translated, using the standard genetic code, to provide an amino acid sequence identical to that translated from the reference nucleic acid sequence.

The term "polypeptide" encompasses both naturally-occurring and non-naturally-occurring proteins and polypeptides, polypeptide fragments and polypeptide mutants, derivatives and analogs. A polypeptide may be monomeric or polymeric. Further, a polypeptide may comprise a number of different modules within a single polypeptide each of which has one or more distinct activities. A preferred polypeptide in accordance with the invention comprises a BSP encoded by a nucleic acid molecule of the instant invention, as well as a fragment, mutant, analog and derivative thereof.

The term "isolated protein" or "isolated polypeptide" is a protein or polypeptide that by virtue of its origin or source of derivation (1) is not associated with naturally associated components that accompany it in its native state, (2) is free of other proteins from the same species (3) is expressed by a cell from a different species, or (4) does not occur in nature. Thus, a polypeptide that is chemically synthesized or synthesized in a cellular system different from the cell from which it naturally originates will be "isolated" from its naturally associated components. A polypeptide or protein may also be rendered substantially free of naturally associated components by isolation, using protein purification techniques well-known in the art.

A protein or polypeptide is "substantially pure," "substantially homogeneous" or "substantially purified" when at least about 60% to 75% of a sample exhibits a single species of polypeptide. The polypeptide or protein may be monomeric or multimeric. A substantially pure polypeptide or protein will typically comprise about 50%, 60%, 70%, 80% or 90% W/W of a protein sample, more usually about 95%, and preferably will be over 99% pure. Protein purity or homogeneity may be indicated by a number of means well-known in the art, such as polyacrylamide gel electrophoresis of a protein sample,

followed by visualizing a single polypeptide band upon staining the gel with a stain well-known in the art. For certain purposes, higher resolution may be provided by using HPLC or other means well-known in the art for purification.

The term "polypeptide fragment" as used herein refers to a polypeptide of the instant invention that has an amino-terminal and/or carboxy-terminal deletion compared to a full-length polypeptide. In a preferred embodiment, the polypeptide fragment is a contiguous sequence in which the amino acid sequence of the fragment is identical to the corresponding positions in the naturally-occurring sequence. Fragments typically are at least 5, 6, 7, 8, 9 or 10 amino acids long, preferably at least 12, 14, 16 or 18 amino acids long, more preferably at least 20 amino acids long, more preferably at least 25, 30, 35, 40 or 45, amino acids, even more preferably at least 50 or 60 amino acids long, and even more preferably at least 70 amino acids long.

A "derivative" refers to polypeptides or fragments thereof that are substantially similar in primary structural sequence but which include, *e.g.*, *in vivo* or *in vitro* chemical and biochemical modifications that are not found in the native polypeptide. Such modifications include, for example, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent cross-links, formation of cystine, formation of pyroglutamate, formylation, gamma-carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination. Other modification include, *e.g.*, labeling with radionuclides, and various enzymatic modifications, as will be readily appreciated by those skilled in the art. A variety of methods for labeling polypeptides and of substituents or labels useful for such purposes are well-known in the art, and include radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H, ligands which bind to labeled antiligands (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands which can serve as specific binding pair members for a labeled ligand. The choice of label depends on the sensitivity required, ease of conjugation with the primer, stability requirements, and available instrumentation. Methods for labeling polypeptides

are well-known in the art. *See* Ausubel (1992), *supra*; Ausubel (1999), *supra*, herein incorporated by reference.

The term "fusion protein" refers to polypeptides of the instant invention comprising polypeptides or fragments coupled to heterologous amino acid sequences. Fusion proteins are useful because they can be constructed to contain two or more desired functional elements from two or more different proteins. A fusion protein comprises at least 10 contiguous amino acids from a polypeptide of interest, more preferably at least 20 or 30 amino acids, even more preferably at least 40, 50 or 60 amino acids, yet more preferably at least 75, 100 or 125 amino acids. Fusion proteins can be produced recombinantly by constructing a nucleic acid sequence which encodes the polypeptide or a fragment thereof in frame with a nucleic acid sequence encoding a different protein or peptide and then expressing the fusion protein. Alternatively, a fusion protein can be produced chemically by crosslinking the polypeptide or a fragment thereof to another protein.

The term "analog" refers to both polypeptide analogs and non-peptide analogs. The term "polypeptide analog" as used herein refers to a polypeptide of the instant invention that is comprised of a segment of at least 25 amino acids that has substantial identity to a portion of an amino acid sequence but which contains non-natural amino acids or non-natural inter-residue bonds. In a preferred embodiment, the analog has the same or similar biological activity as the native polypeptide. Typically, polypeptide analogs comprise a conservative amino acid substitution (or insertion or deletion) with respect to the naturally-occurring sequence. Analogs typically are at least 20 amino acids long, preferably at least 50 amino acids long or longer, and can often be as long as a full-length naturally-occurring polypeptide.

The term "non-peptide analog" refers to a compound with properties that are analogous to those of a reference polypeptide of the instant invention. A non-peptide compound may also be termed a "peptide mimetic" or a "peptidomimetic." Such compounds are often developed with the aid of computerized molecular modeling. Peptide mimetics that are structurally similar to useful peptides may be used to produce an equivalent effect. Generally, peptidomimetics are structurally similar to a paradigm polypeptide (*i.e.*, a polypeptide that has a desired biochemical property or pharmacological activity), but have one or more peptide linkages optionally replaced by a linkage selected from the group consisting of: $--CH_2NH--$, $--CH_2S--$, $--CH_2-CH_2--$, $--CH=CH--$(cis and trans), $--COCH_2--$, $--CH(OH)CH_2--$, and $--CH_2SO--$, by methods

well-known in the art. Systematic substitution of one or more amino acids of a consensus sequence with a D-amino acid of the same type (*e.g.*, D-lysine in place of L-lysine) may also be used to generate more stable peptides. In addition, constrained peptides comprising a consensus sequence or a substantially identical consensus sequence variation may be generated by methods known in the art (Rizo *et al.*, *Ann. Rev. Biochem.* 61:387-418 (1992), incorporated herein by reference). For example, one may add internal cysteine residues capable of forming intramolecular disulfide bridges which cyclize the peptide.

A "polypeptide mutant" or "mutein" refers to a polypeptide of the instant invention whose sequence contains substitutions, insertions or deletions of one or more amino acids compared to the amino acid sequence of a native or wild-type protein. A mutein may have one or more amino acid point substitutions, in which a single amino acid at a position has been changed to another amino acid, one or more insertions and/or deletions, in which one or more amino acids are inserted or deleted, respectively, in the sequence of the naturally-occurring protein, and/or truncations of the amino acid sequence at either or both the amino or carboxy termini. Further, a mutein may have the same or different biological activity as the naturally-occurring protein. For instance, a mutein may have an increased or decreased biological activity. A mutein has at least 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are muteins having 80%, 85% or 90% sequence similarity to the wild type protein. In an even more preferred embodiment, a mutein exhibits 95% sequence identity, even more preferably 97%, even more preferably 98% and even more preferably 99%. Sequence similarity may be measured by any common sequence analysis algorithm, such as Gap or Bestfit.

Preferred amino acid substitutions are those which: (1) reduce susceptibility to proteolysis, (2) reduce susceptibility to oxidation, (3) alter binding affinity for forming protein complexes, (4) alter binding affinity or enzymatic activity, and (5) confer or modify other physicochemical or functional properties of such analogs. For example, single or multiple amino acid substitutions (preferably conservative amino acid substitutions) may be made in the naturally-occurring sequence (preferably in the portion of the polypeptide outside the domain(s) forming intermolecular contacts. In a preferred embodiment, the amino acid substitutions are moderately conservative substitutions or conservative substitutions. In a more preferred embodiment, the amino acid substitutions are conservative substitutions. A conservative amino acid substitution should not

substantially change the structural characteristics of the parent sequence (*e.g.*, a replacement amino acid should not tend to disrupt a helix that occurs in the parent sequence, or disrupt other types of secondary structure that characterizes the parent sequence). Examples of art-recognized polypeptide secondary and tertiary structures are described in Creighton (ed.), Proteins, Structures and Molecular Principles, W. H. Freeman and Company (1984); Branden *et al.* (ed.), Introduction to Protein Structure, Garland Publishing (1991); Thornton *et al.*, *Nature* 354:105-106 (1991), each of which are incorporated herein by reference.

As used herein, the twenty conventional amino acids and their abbreviations follow conventional usage. *See* Golub *et al.* (eds.), Immunology - A Synthesis 2nd Ed., Sinauer Associates (1991), which is incorporated herein by reference. Stereoisomers (*e.g.*, D-amino acids) of the twenty conventional amino acids, unnatural amino acids such as α-, α-disubstituted amino acids, N-alkyl amino acids, and other unconventional amino acids may also be suitable components for polypeptides of the present invention. Examples of unconventional amino acids include: 4-hydroxyproline, γ-carboxyglutamate, ε-N,N,N-trimethyllysine, ε-N-acetyllysine, O-phosphoserine, N-acetylserine, N-formylmethionine, 3-methylhistidine, 5-hydroxylysine, s-N-methylarginine, and other similar amino acids and imino acids (*e.g.*, 4-hydroxyproline). In the polypeptide notation used herein, the lefthand direction is the amino terminal direction and the right hand direction is the carboxy-terminal direction, in accordance with standard usage and convention.

A protein has "homology" or is "homologous" to a protein from another organism if the encoded amino acid sequence of the protein has a similar sequence to the encoded amino acid sequence of a protein of a different organism and has a similar biological activity or function. Alternatively, a protein may have homology or be homologous to another protein if the two proteins have similar amino acid sequences and have similar biological activities or functions. Although two proteins are said to be "homologous," this does not imply that there is necessarily an evolutionary relationship between the proteins. Instead, the term "homologous" is defined to mean that the two proteins have similar amino acid sequences and similar biological activities or functions. In a preferred embodiment, a homologous protein is one that exhibits 50% sequence similarity to the wild type protein, preferred is 60% sequence similarity, more preferred is 70% sequence similarity. Even more preferred are homologous proteins that exhibit 80%, 85% or 90%

sequence similarity to the wild type protein. In a yet more preferred embodiment, a homologous protein exhibits 95%, 97%, 98% or 99% sequence similarity.

When "sequence similarity" is used in reference to proteins or peptides, it is recognized that residue positions that are not identical often differ by conservative amino acid substitutions. In a preferred embodiment, a polypeptide that has "sequence similarity" comprises conservative or moderately conservative amino acid substitutions. A "conservative amino acid substitution" is one in which an amino acid residue is substituted by another amino acid residue having a side chain (R group) with similar chemical properties (*e.g.*, charge or hydrophobicity). In general, a conservative amino acid substitution will not substantially change the functional properties of a protein. In cases where two or more amino acid sequences differ from each other by conservative substitutions, the percent sequence identity or degree of similarity may be adjusted upwards to correct for the conservative nature of the substitution. Means for making this adjustment are well-known to those of skill in the art. *See, e.g.*, Pearson, *Methods Mol. Biol.* 24: 307-31 (1994), herein incorporated by reference.

For instance, the following six groups each contain amino acids that are conservative substitutions for one another:

1) Serine (S), Threonine (T);
2) Aspartic Acid (D), Glutamic Acid (E);
3) Asparagine (N), Glutamine (Q);
4) Arginine (R), Lysine (K);
5) Isoleucine (I), Leucine (L), Methionine (M), Alanine (A), Valine (V), and
6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

Alternatively, a conservative replacement is any change having a positive value in the PAM250 log-likelihood matrix disclosed in Gonnet *et al.*, *Science* 256: 1443-45 (1992), herein incorporated by reference. A "moderately conservative" replacement is any change having a nonnegative value in the PAM250 log-likelihood matrix.

Sequence similarity for polypeptides, which is also referred to as sequence identity, is typically measured using sequence analysis software. Protein analysis software matches similar sequences using measures of similarity assigned to various substitutions, deletions and other modifications, including conservative amino acid substitutions. For instance, GCG contains programs such as "Gap" and "Bestfit" which can be used with default parameters to determine sequence homology or sequence identity between closely related polypeptides, such as homologous polypeptides from

different species of organisms or between a wild type protein and a mutein thereof. *See, e.g.*, GCG Version 6.1. Other programs include FASTA, discussed *supra*.

A preferred algorithm when comparing a sequence of the invention to a database containing a large number of sequences from different organisms is the computer program BLAST, especially blastp or tblastn. *See, e.g.*, Altschul *et al.*, *J. Mol. Biol.* 215: 403-410 (1990); Altschul *et al.*, *Nucleic Acids Res.* 25:3389-402 (1997); herein incorporated by reference. Preferred parameters for blastp are:

Expectation value:	10 (default)
Filter:	seg (default)
Cost to open a gap:	11 (default)
Cost to extend a gap:	1 (default
Max. alignments:	100 (default)
Word size:	11 (default)
No. of descriptions:	100 (default)
Penalty Matrix:	BLOSUM62

The length of polypeptide sequences compared for homology will generally be at least about 16 amino acid residues, usually at least about 20 residues, more usually at least about 24 residues, typically at least about 28 residues, and preferably more than about 35 residues. When searching a database containing sequences from a large number of different organisms, it is preferable to compare amino acid sequences.

Database searching using amino acid sequences can be measured by algorithms other than blastp are known in the art. For instance, polypeptide sequences can be compared using FASTA, a program in GCG Version 6.1. FASTA (*e.g.*, FASTA2 and FASTA3) provides alignments and percent sequence identity of the regions of the best overlap between the query and search sequences (Pearson (1990), *supra*; Pearson (2000), *supra*. For example, percent sequence identity between amino acid sequences can be determined using FASTA with its default or recommended parameters (a word size of 2 and the PAM250 scoring matrix), as provided in GCG Version 6.1, herein incorporated by reference.

An "antibody" refers to an intact immunoglobulin, or to an antigen-binding portion thereof that competes with the intact antibody for specific binding to a molecular species, *e.g.*, a polypeptide of the instant invention. Antigen-binding portions may be produced by recombinant DNA techniques or by enzymatic or chemical cleavage of intact antibodies. Antigen-binding portions include, *inter alia*, Fab, Fab', $F(ab')_2$, Fv,

dAb, and complementarity determining region (CDR) fragments, single-chain antibodies (scFv), chimeric antibodies, diabodies and polypeptides that contain at least a portion of an immunoglobulin that is sufficient to confer specific antigen binding to the polypeptide. An Fab fragment is a monovalent fragment consisting of the VL, VH, CL and CH1 domains; an $F(ab')_2$ fragment is a bivalent fragment comprising two Fab fragments linked by a disulfide bridge at the hinge region; an Fd fragment consists of the VH and CH1 domains; an Fv fragment consists of the VL and VH domains of a single arm of an antibody; and a dAb fragment consists of a VH domain. *See, e.g.,* Ward *et al.*, *Nature* 341: 544-546 (1989).

By "bind specifically" and "specific binding" is here intended the ability of the antibody to bind to a first molecular species in preference to binding to other molecular species with which the antibody and first molecular species are admixed. An antibody is said specifically to "recognize" a first molecular species when it can bind specifically to that first molecular species.

A single-chain antibody (scFv) is an antibody in which a VL and VH region are paired to form a monovalent molecule via a synthetic linker that enables them to be made as a single protein chain. *See, e.g.,* Bird *et al.*, *Science* 242: 423-426 (1988); Huston *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5879-5883 (1988). Diabodies are bivalent, bispecific antibodies in which VH and VL domains are expressed on a single polypeptide chain, but using a linker that is too short to allow for pairing between the two domains on the same chain, thereby forcing the domains to pair with complementary domains of another chain and creating two antigen binding sites. *See e.g.,* Holliger *et al.*, *Proc. Natl. Acad. Sci. USA* 90: 6444-6448 (1993); Poljak *et al.*, *Structure* 2: 1121-1123 (1994). One or more CDRs may be incorporated into a molecule either covalently or noncovalently to make it an immunoadhesin. An immunoadhesin may incorporate the CDR(s) as part of a larger polypeptide chain, may covalently link the CDR(s) to another polypeptide chain, or may incorporate the CDR(s) noncovalently. The CDRs permit the immunoadhesin to specifically bind to a particular antigen of interest. A chimeric antibody is an antibody that contains one or more regions from one antibody and one or more regions from one or more other antibodies.

An antibody may have one or more binding sites. If there is more than one binding site, the binding sites may be identical to one another or may be different. For instance, a naturally-occurring immunoglobulin has two identical binding sites, a single-

chain antibody or Fab fragment has one binding site, while a "bispecific" or "bifunctional" antibody has two different binding sites.

An "isolated antibody" is an antibody that (1) is not associated with naturally-associated components, including other naturally-associated antibodies, that accompany it in its native state, (2) is free of other proteins from the same species, (3) is expressed by a cell from a different species, or (4) does not occur in nature. It is known that purified proteins, including purified antibodies, may be stabilized with non-naturally-associated components. The non-naturally-associated component may be a protein, such as albumin (*e.g.*, BSA) or a chemical such as polyethylene glycol (PEG).

A "neutralizing antibody" or "an inhibitory antibody" is an antibody that inhibits the activity of a polypeptide or blocks the binding of a polypeptide to a ligand that normally binds to it. An "activating antibody" is an antibody that increases the activity of a polypeptide.

The term "epitope" includes any protein determinant capable of specifically binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three-dimensional structural characteristics, as well as specific charge characteristics. An antibody is said to specifically bind an antigen when the dissociation constant is less than1 μM, preferably less than100 nM and most preferably less than 10 nM.

The term "patient" as used herein includes human and veterinary subjects.

Throughout this specification and claims, the word "comprise," or variations such as "comprises" or "comprising," will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

The term "breast specific" refers to a nucleic acid molecule or polypeptide that is expressed predominantly in the breast as compared to other tissues in the body. In a preferred embodiment, a "breast specific" nucleic acid molecule or polypeptide is expressed at a level that is 5-fold higher than any other tissue in the body. In a more preferred embodiment, the "breast specific" nucleic acid molecule or polypeptide is expressed at a level that is 10-fold higher than any other tissue in the body, more preferably at least 15-fold, 20-fold, 25-fold, 50-fold or 100-fold higher than any other tissue in the body. Nucleic acid molecule levels may be measured by nucleic acid hybridization, such as Northern blot hybridization, or quantitative PCR. Polypeptide

levels may be measured by any method known to accurately quantitate protein levels, such as Western blot analysis.

Nucleic Acid Molecules, Regulatory Sequences, Vectors, Host Cells and Recombinant Methods of Making Polypeptides

Nucleic Acid Molecules

One aspect of the invention provides isolated nucleic acid molecules that are specific to the breast or to breast cells or tissue or that are derived from such nucleic acid molecules. These isolated breast specific nucleic acids (BSNAs) may comprise a cDNA, a genomic DNA, RNA, or a fragment of one of these nucleic acids, or may be a non-naturally-occurring nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule encodes a polypeptide that is specific to breast, a breast-specific polypeptide (BSP). In a more preferred embodiment, the nucleic acid molecule encodes a polypeptide that comprises an amino acid sequence of SEQ ID NO: 172 through 295. In another highly preferred embodiment, the nucleic acid molecule comprises a nucleic acid sequence of SEQ ID NO: 1 through 171.

A BSNA may be derived from a human or from another animal. In a preferred embodiment, the BSNA is derived from a human or other mammal. In a more preferred embodiment, the BSNA is derived from a human or other primate. In an even more preferred embodiment, the BSNA is derived from a human.

By "nucleic acid molecule" for purposes of the present invention, it is also meant to be inclusive of nucleic acid sequences that selectively hybridize to a nucleic acid molecule encoding a BSNA or a complement thereof. The hybridizing nucleic acid molecule may or may not encode a polypeptide or may not encode a BSP. However, in a preferred embodiment, the hybridizing nucleic acid molecule encodes a BSP. In a more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule that encodes a polypeptide comprising an amino acid sequence of SEQ ID NO: 172 through 295. In an even more preferred embodiment, the invention provides a nucleic acid molecule that selectively hybridizes to a nucleic acid molecule comprising the nucleic acid sequence of SEQ ID NO: 1 through 171.

In a preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a BSP under low stringency conditions. In a more preferred embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a BSP under moderate stringency conditions. In a more preferred

embodiment, the nucleic acid molecule selectively hybridizes to a nucleic acid molecule encoding a BSP under high stringency conditions. In an even more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence of SEQ ID NO: 172 through 295. In a yet more preferred embodiment, the nucleic acid molecule hybridizes under low, moderate or high stringency conditions to a nucleic acid molecule comprising a nucleic acid sequence selected from SEQ ID NO: 1 through 171. In a preferred embodiment of the invention, the hybridizing nucleic acid molecule may be used to express recombinantly a polypeptide of the invention.

By "nucleic acid molecule" as used herein it is also meant to be inclusive of sequences that exhibits substantial sequence similarity to a nucleic acid encoding a BSP or a complement of the encoding nucleic acid molecule. In a preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding human BSP. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 172 through 295. In a preferred embodiment, the similar nucleic acid molecule is one that has at least 60% sequence identity with a nucleic acid molecule encoding a BSP, such as a polypeptide having an amino acid sequence of SEQ ID NO: 172 through 295, more preferably at least 70%, even more preferably at least 80% and even more preferably at least 85%. In a more preferred embodiment, the similar nucleic acid molecule is one that has at least 90% sequence identity with a nucleic acid molecule encoding a BSP, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a nucleic acid molecule encoding a BSP.

In another preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a BSNA or its complement. In a more preferred embodiment, the nucleic acid molecule exhibits substantial sequence similarity to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 171. In a preferred embodiment, the nucleic acid molecule is one that has at least 60% sequence identity with a BSNA, such as one having a nucleic acid sequence of SEQ ID NO: 1 through 171, more preferably at least 70%, even more preferably at least 80% and even more

preferably at least 85%. In a more preferred embodiment, the nucleic acid molecule is one that has at least 90% sequence identity with a BSNA, more preferably at least 95%, more preferably at least 97%, even more preferably at least 98%, and still more preferably at least 99%. In another highly preferred embodiment, the nucleic acid molecule is one that has at least 99.5%, 99.6%, 99.7%, 99.8% or 99.9% sequence identity with a BSNA.

A nucleic acid molecule that exhibits substantial sequence similarity may be one that exhibits sequence identity over its entire length to a BSNA or to a nucleic acid molecule encoding a BSP, or may be one that is similar over only a part of its length. In this case, the part is at least 50 nucleotides of the BSNA or the nucleic acid molecule encoding a BSP, preferably at least 100 nucleotides, more preferably at least 150 or 200 nucleotides, even more preferably at least 250 or 300 nucleotides, still more preferably at least 400 or 500 nucleotides.

The substantially similar nucleic acid molecule may be a naturally-occurring one that is derived from another species, especially one derived from another primate, wherein the similar nucleic acid molecule encodes an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 172 through 295 or demonstrates significant sequence identity to the nucleotide sequence of SEQ ID NO: 1 through 171. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule from a human, when the BSNA is a member of a gene family. The similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, hamster, cow, horse and pig; and wild animals, *e.g.*, monkey, fox, lions, tigers, bears, giraffes, zebras, etc. The substantially similar nucleic acid molecule may also be a naturally-occurring nucleic acid molecule derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring substantially similar nucleic acid molecule may be isolated directly from humans or other species. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by random mutation of a nucleic acid molecule. In another embodiment, the substantially similar nucleic acid molecule may be one that is experimentally produced by directed mutation of a BSNA. Further, the substantially similar nucleic acid molecule may or may not be a BSNA. However, in a preferred embodiment, the substantially similar nucleic acid molecule is a BSNA.

By "nucleic acid molecule" it is also meant to be inclusive of allelic variants of a BSNA or a nucleic acid encoding a BSP. For instance, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes. In fact, more than 1.4 million SNPs have already identified in the human genome, International Human Genome Sequencing Consortium, *Nature* 409: 860-921 (2001). Thus, the sequence determined from one individual of a species may differ from other allelic forms present within the population. Additionally, small deletions and insertions, rather than single nucleotide polymorphisms, are not uncommon in the general population, and often do not alter the function of the protein. Further, amino acid substitutions occur frequently among natural allelic variants, and often do not substantially change protein function.

In a preferred embodiment, the nucleic acid molecule comprising an allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that encodes a BSP. In a more preferred embodiment, the gene is transcribed into an mRNA that encodes a BSP comprising an amino acid sequence of SEQ ID NO: 172 through 295. In another preferred embodiment, the allelic variant is a variant of a gene, wherein the gene is transcribed into an mRNA that is a BSNA. In a more preferred embodiment, the gene is transcribed into an mRNA that comprises the nucleic acid sequence of SEQ ID NO: 1 through 171. In a preferred embodiment, the allelic variant is a naturally-occurring allelic variant in the species of interest. In a more preferred embodiment, the species of interest is human.

By "nucleic acid molecule" it is also meant to be inclusive of a part of a nucleic acid sequence of the instant invention. The part may or may not encode a polypeptide, and may or may not encode a polypeptide that is a BSP. However, in a preferred embodiment, the part encodes a BSP. In one aspect, the invention comprises a part of a BSNA. In a second aspect, the invention comprises a part of a nucleic acid molecule that hybridizes or exhibits substantial sequence similarity to a BSNA. In a third aspect, the invention comprises a part of a nucleic acid molecule that is an allelic variant of a BSNA. In a fourth aspect, the invention comprises a part of a nucleic acid molecule that encodes a BSP. A part comprises at least 10 nucleotides, more preferably at least 15, 17, 18, 20, 25, 30, 35, 40, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400 or 500 nucleotides. The maximum size of a nucleic acid part is one nucleotide shorter than the sequence of the nucleic acid molecule encoding the full-length protein.

By "nucleic acid molecule" it is also meant to be inclusive of sequence that encoding a fusion protein, a homologous protein, a polypeptide fragment, a mutein or a polypeptide analog, as described below.

Nucleotide sequences of the instantly-described nucleic acids were determined by sequencing a DNA molecule that had resulted, directly or indirectly, from at least one enzymatic polymerization reaction (*e.g.*, reverse transcription and/or polymerase chain reaction) using an automated sequencer (such as the MegaBACE™ 1000, Molecular Dynamics, Sunnyvale, CA, USA). Further, all amino acid sequences of the polypeptides of the present invention were predicted by translation from the nucleic acid sequences so determined, unless otherwise specified.

In a preferred embodiment of the invention, the nucleic acid molecule contains modifications of the native nucleic acid molecule. These modifications include nonnative internucleoside bonds, post-synthetic modifications or altered nucleotide analogues. One having ordinary skill in the art would recognize that the type of modification that can be made will depend upon the intended use of the nucleic acid molecule. For instance, when the nucleic acid molecule is used as a hybridization probe, the range of such modifications will be limited to those that permit sequence-discriminating base pairing of the resulting nucleic acid. When used to direct expression of RNA or protein *in vitro* or *in vivo*, the range of such modifications will be limited to those that permit the nucleic acid to function properly as a polymerization substrate. When the isolated nucleic acid is used as a therapeutic agent, the modifications will be limited to those that do not confer toxicity upon the isolated nucleic acid.

In a preferred embodiment, isolated nucleic acid molecules can include nucleotide analogues that incorporate labels that are directly detectable, such as radiolabels or fluorophores, or nucleotide analogues that incorporate labels that can be visualized in a subsequent reaction, such as biotin or various haptens. In a more preferred embodiment, the labeled nucleic acid molecule may be used as a hybridization probe.

Common radiolabeled analogues include those labeled with ^{33}P, ^{32}P, and ^{35}S, such as α-^{32}P-dATP, α-^{32}P-dCTP, α-^{32}P-dGTP, α-^{32}P-dTTP, α-^{32}P-3'dATP, α-^{32}P-ATP, α-^{32}P-CTP, α-^{32}P-GTP, α-^{32}P-UTP, α-^{35}S-dATP, α-^{35}S-GTP, α-^{33}P-dATP, and the like.

Commercially available fluorescent nucleotide analogues readily incorporated into the nucleic acids of the present invention include Cy3-dCTP, Cy3-dUTP, Cy5-dCTP, Cy3-dUTP (Amersham Pharmacia Biotech, Piscataway, New Jersey, USA), fluorescein-12-dUTP, tetramethylrhodamine-6-dUTP, Texas Red®-5-dUTP, Cascade

Blue®-7-dUTP, BODIPY® FL-14-dUTP, BODIPY® TMR-14-dUTP, BODIPY® TR-14-dUTP, Rhodamine Green™-5-dUTP, Oregon Green® 488-5-dUTP, Texas Red®-12-dUTP, BODIPY® 630/650-14-dUTP, BODIPY® 650/665-14-dUTP, Alexa Fluor® 488-5-dUTP, Alexa Fluor® 532-5-dUTP, Alexa Fluor® 568-5-dUTP, Alexa Fluor® 594-5-dUTP, Alexa Fluor® 546-14-dUTP, fluorescein-12-UTP, tetramethylrhodamine-6-UTP, Texas Red®-5-UTP, Cascade Blue®-7-UTP, BODIPY® FL-14-UTP, BODIPY® TMR-14-UTP, BODIPY® TR-14-UTP, Rhodamine Green™-5-UTP, Alexa Fluor® 488-5-UTP, Alexa Fluor® 546-14-UTP (Molecular Probes, Inc. Eugene, OR, USA). One may also custom synthesize nucleotides having other fluorophores. *See* Henegariu *et al.*, *Nature Biotechnol.* 18: 345-348 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Haptens that are commonly conjugated to nucleotides for subsequent labeling include biotin (biotin-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA; biotin-21-UTP, biotin-21-dUTP, Clontech Laboratories, Inc., Palo Alto, CA, USA), digoxigenin (DIG-11-dUTP, alkali labile, DIG-11-UTP, Roche Diagnostics Corp., Indianapolis, IN, USA), and dinitrophenyl (dinitrophenyl-11-dUTP, Molecular Probes, Inc., Eugene, OR, USA).

Nucleic acid molecules can be labeled by incorporation of labeled nucleotide analogues into the nucleic acid. Such analogues can be incorporated by enzymatic polymerization, such as by nick translation, random priming, polymerase chain reaction (PCR), terminal transferase tailing, and end-filling of overhangs, for DNA molecules, and *in vitro* transcription driven, *e.g.*, from phage promoters, such as T7, T3, and SP6, for RNA molecules. Commercial kits are readily available for each such labeling approach. Analogues can also be incorporated during automated solid phase chemical synthesis. Labels can also be incorporated after nucleic acid synthesis, with the 5' phosphate and 3' hydroxyl providing convenient sites for post-synthetic covalent attachment of detectable labels.

Other post-synthetic approaches also permit internal labeling of nucleic acids. For example, fluorophores can be attached using a cisplatin reagent that reacts with the N7 of guanine residues (and, to a lesser extent, adenine bases) in DNA, RNA, and PNA to provide a stable coordination complex between the nucleic acid and fluorophore label (Universal Linkage System) (available from Molecular Probes, Inc., Eugene, OR, USA and Amersham Pharmacia Biotech, Piscataway, NJ, USA); *see* Alers *et al.*, *Genes, Chromosomes & Cancer* 25: 301- 305 (1999); Jelsma *et al.*, *J. NIH Res.* 5: 82 (1994);

Van Belkum *et al.*, *BioTechniques* 16: 148-153 (1994), incorporated herein by reference. As another example, nucleic acids can be labeled using a disulfide-containing linker (FastTag™ Reagent, Vector Laboratories, Inc., Burlingame, CA, USA) that is photo- or thermally-coupled to the target nucleic acid using aryl azide chemistry; after reduction, a free thiol is available for coupling to a hapten, fluorophore, sugar, affinity ligand, or other marker.

One or more independent or interacting labels can be incorporated into the nucleic acid molecules of the present invention. For example, both a fluorophore and a moiety that in proximity thereto acts to quench fluorescence can be included to report specific hybridization through release of fluorescence quenching or to report exonucleotidic excision. *See, e.g.,* Tyagi *et al.*, *Nature Biotechnol.* 14: 303-308 (1996); Tyagi *et al.*, *Nature Biotechnol.* 16: 49-53 (1998); Sokol *et al.*, *Proc. Natl. Acad. Sci. USA* 95: 11538-11543 (1998); Kostrikis *et al.*, *Science* 279: 1228-1229 (1998); Marras *et al.*, *Genet. Anal.* 14: 151-156 (1999); U. S. Patent 5,846,726; 5,925,517; 5,925,517; 5,723,591 and 5,538,848; Holland *et al.*, *Proc. Natl. Acad. Sci. USA* 88: 7276-7280 (1991); Heid *et al.*, *Genome Res.* 6(10): 986-94 (1996); Kuimelis *et al.*, *Nucleic Acids Symp. Ser.* (37): 255-6 (1997); the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules of the invention may be modified by altering one or more native phosphodiester internucleoside bonds to more nuclease-resistant, internucleoside bonds. *See* Hartmann *et al.* (eds.), Manual of Antisense Methodology: Perspectives in Antisense Science, Kluwer Law International (1999); Stein *et al.* (eds.), Applied Antisense Oligonucleotide Technology, Wiley-Liss (1998); Chadwick *et al.* (eds.), Oligonucleotides as Therapeutic Agents - Symposium No. 209, John Wiley & Son Ltd (1997); the disclosures of which are incorporated herein by reference in their entireties. Such altered internucleoside bonds are often desired for antisense techniques or for targeted gene correction. *See* Gamper *et al.*, *Nucl. Acids Res.* 28(21): 4332-4339 (2000), the disclosure of which is incorporated herein by reference in its entirety.

Modified oligonucleotide backbones include, without limitation, phosphorothioates, chiral phosphorothioates, phosphorodithioates, phosphotriesters, aminoalkylphosphotriesters, methyl and other alkyl phosphonates including 3'-alkylene phosphonates and chiral phosphonates, phosphinates, phosphoramidates including 3'-amino phosphoramidate and aminoalkylphosphoramidates, thionophosphoramidates, thionoalkylphosphonates, thionoalkylphosphotriesters, and boranophosphates having

normal 3'-5' linkages, 2'-5' linked analogs of these, and those having inverted polarity wherein the adjacent pairs of nucleoside units are linked 3'-5' to 5'-3' or 2'-5' to 5'-2'. Representative United States patents that teach the preparation of the above phosphorus-containing linkages include, but are not limited to, U. S. Patents 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; and 5,625,050, the disclosures of which are incorporated herein by reference in their entireties. In a preferred embodiment, the modified internucleoside linkages may be used for antisense techniques.

Other modified oligonucleotide backbones do not include a phosphorus atom, but have backbones that are formed by short chain alkyl or cycloalkyl internucleoside linkages, mixed heteroatom and alkyl or cycloalkyl internucleoside linkages, or one or more short chain heteroatomic or heterocyclic internucleoside linkages. These include those having morpholino linkages (formed in part from the sugar portion of a nucleoside); siloxane backbones; sulfide, sulfoxide and sulfone backbones; formacetyl and thioformacetyl backbones; methylene formacetyl and thioformacetyl backbones; alkene containing backbones; sulfamate backbones; methyleneimino and methylenehydrazino backbones; sulfonate and sulfonamide backbones; amide backbones; and others having mixed N, O, S and CH_2 component parts. Representative U.S. patents that teach the preparation of the above backbones include, but are not limited to, U.S. Patent 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437 and 5,677,439; the disclosures of which are incorporated herein by reference in their entireties.

In other preferred oligonucleotide mimetics, both the sugar and the internucleoside linkage are replaced with novel groups, such as peptide nucleic acids (PNA). In PNA compounds, the phosphodiester backbone of the nucleic acid is replaced with an amide-containing backbone, in particular by repeating N-(2-aminoethyl) glycine units linked by amide bonds. Nucleobases are bound directly or indirectly to aza nitrogen atoms of the amide portion of the backbone, typically by methylene carbonyl linkages. PNA can be synthesized using a modified peptide synthesis protocol. PNA oligomers can be synthesized by both Fmoc and tBoc methods. Representative U.S.

patents that teach the preparation of PNA compounds include, but are not limited to, U.S Patent 5,539,082; 5,714,331; and 5,719,262, each of which is herein incorporated by reference. Automated PNA synthesis is readily achievable on commercial synthesizers (*see, e.g.*, "PNA User's Guide," Rev. 2, February 1998, Perseptive Biosystems Part No. 60138, Applied Biosystems, Inc., Foster City, CA).

PNA molecules are advantageous for a number of reasons. First, because the PNA backbone is uncharged, PNA/DNA and PNA/RNA duplexes have a higher thermal stability than is found in DNA/DNA and DNA/RNA duplexes. The Tm of a PNA/DNA or PNA/RNA duplex is generally 1°C higher per base pair than the Tm of the corresponding DNA/DNA or DNA/RNA duplex (in 100 mM NaCl). Second, PNA molecules can also form stable PNA/DNA complexes at low ionic strength, under conditions in which DNA/DNA duplex formation does not occur. Third, PNA also demonstrates greater specificity in binding to complementary DNA because a PNA/DNA mismatch is more destabilizing than DNA/DNA mismatch. A single mismatch in mixed a PNA/DNA 15-mer lowers the Tm by 8–20°C (15°C on average). In the corresponding DNA/DNA duplexes, a single mismatch lowers the Tm by 4–16°C (11°C on average). Because PNA probes can be significantly shorter than DNA probes, their specificity is greater. Fourth, PNA oligomers are resistant to degradation by enzymes, and the lifetime of these compounds is extended both *in vivo* and *in vitro* because nucleases and proteases do not recognize the PNA polyamide backbone with nucleobase sidechains. *See, e.g.*, Ray *et al.*, *FASEB J.* 14(9): 1041-60 (2000); Nielsen *et al.*, *Pharmacol Toxicol.* 86(1): 3-7 (2000); Larsen *et al.*, *Biochim Biophys Acta.* 1489(1): 159-66 (1999); Nielsen, *Curr. Opin. Struct. Biol.* 9(3): 353-7 (1999), and Nielsen, *Curr. Opin. Biotechnol.* 10(1): 71-5 (1999), the disclosures of which are incorporated herein by reference in their entireties.

Nucleic acid molecules may be modified compared to their native structure throughout the length of the nucleic acid molecule or can be localized to discrete portions thereof. As an example of the latter, chimeric nucleic acids can be synthesized that have discrete DNA and RNA domains and that can be used for targeted gene repair and modified PCR reactions, as further described in U.S. Patents 5,760,012 and 5,731,181, Misra *et al.*, *Biochem.* 37: 1917-1925 (1998); and Finn *et al.*, *Nucl. Acids Res.* 24: 3357-3363 (1996), the disclosures of which are incorporated herein by reference in their entireties.

Unless otherwise specified, nucleic acids of the present invention can include any topological conformation appropriate to the desired use; the term thus explicitly

comprehends, among others, single-stranded, double-stranded, triplexed, quadruplexed, partially double-stranded, partially-triplexed, partially-quadruplexed, branched, hairpinned, circular, and padlocked conformations. Padlock conformations and their utilities are further described in Banér *et al.*, *Curr. Opin. Biotechnol.* 12: 11-15 (2001); Escude *et al.*, *Proc. Natl. Acad. Sci. USA* 14: 96(19):10603-7 (1999); Nilsson *et al.*, *Science* 265(5181): 2085-8 (1994), the disclosures of which are incorporated herein by reference in their entireties. Triplex and quadruplex conformations, and their utilities, are reviewed in Praseuth *et al.*, *Biochim. Biophys. Acta.* 1489(1): 181-206 (1999); Fox, *Curr. Med. Chem.* 7(1): 17-37 (2000); Kochetkova *et al.*, *Methods Mol. Biol.* 130: 189-201 (2000); Chan *et al.*, *J. Mol. Med.* 75(4): 267-82 (1997), the disclosures of which are incorporated herein by reference in their entireties.

Methods for Using Nucleic Acid Molecules as Probes and Primers

The isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize, and quantify hybridizing nucleic acids in, and isolate hybridizing nucleic acids from, both genomic and transcript-derived nucleic acid samples. When free in solution, such probes are typically, but not invariably, detectably labeled; bound to a substrate, as in a microarray, such probes are typically, but not invariably unlabeled.

In one embodiment, the isolated nucleic acids of the present invention can be used as probes to detect and characterize gross alterations in the gene of a BSNA, such as deletions, insertions, translocations, and duplications of the BSNA genomic locus through fluorescence *in situ* hybridization (FISH) to chromosome spreads. *See, e.g.*, Andreeff *et al.* (eds.), Introduction to Fluorescence *In Situ* Hybridization: Principles and Clinical Applications, John Wiley & Sons (1999), the disclosure of which is incorporated herein by reference in its entirety. The isolated nucleic acids of the present invention can be used as probes to assess smaller genomic alterations using, *e.g.*, Southern blot detection of restriction fragment length polymorphisms. The isolated nucleic acid molecules of the present invention can be used as probes to isolate genomic clones that include the nucleic acid molecules of the present invention, which thereafter can be restriction mapped and sequenced to identify deletions, insertions, translocations, and substitutions (single nucleotide polymorphisms, SNPs) at the sequence level.

In another embodiment, the isolated nucleic acid molecules of the present invention can be used as probes to detect, characterize, and quantify BSNA in, and

isolate BSNA from, transcript-derived nucleic acid samples. In one aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by length, and quantify mRNA by Northern blot of total or poly-A^+-selected RNA samples. In another aspect, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to detect, characterize by location, and quantify mRNA by *in situ* hybridization to tissue sections. *See, e.g.*, Schwarchzacher *et al.*, In Situ Hybridization, Springer-Verlag New York (2000), the disclosure of which is incorporated herein by reference in its entirety. In another preferred embodiment, the isolated nucleic acid molecules of the present invention can be used as hybridization probes to measure the representation of clones in a cDNA library or to isolate hybridizing nucleic acid molecules acids from cDNA libraries, permitting sequence level characterization of mRNAs that hybridize to BSNAs, including, without limitations, identification of deletions, insertions, substitutions, truncations, alternatively spliced forms and single nucleotide polymorphisms. In yet another preferred embodiment, the nucleic acid molecules of the instant invention may be used in microarrays.

All of the aforementioned probe techniques are well within the skill in the art, and are described at greater length in standard texts such as Sambrook (2001), *supra*; Ausubel (1999), *supra*; and Walker *et al.* (eds.), The Nucleic Acids Protocols Handbook, Humana Press (2000), the disclosures of which are incorporated herein by reference in their entirety.

Thus, in one embodiment, a nucleic acid molecule of the invention may be used as a probe or primer to identify or amplify a second nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of the invention. In a preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a BSP. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule encoding a polypeptide having an amino acid sequence of SEQ ID NO: 172 through 295. In another preferred embodiment, the probe or primer is derived from a BSNA. In a more preferred embodiment, the probe or primer is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 171.

In general, a probe or primer is at least 10 nucleotides in length, more preferably at least 12, more preferably at least 14 and even more preferably at least 16 or 17 nucleotides in length. In an even more preferred embodiment, the probe or primer is at least 18 nucleotides in length, even more preferably at least 20 nucleotides and even more preferably at least 22 nucleotides in length. Primers and probes may also be longer

in length. For instance, a probe or primer may be 25 nucleotides in length, or may be 30, 40 or 50 nucleotides in length. Methods of performing nucleic acid hybridization using oligonucleotide probes are well-known in the art. *See, e.g.*, Sambrook *et al.*, 1989, *supra*, Chapter 11 and pp. 11.31-11.32 and 11.40-11.44, which describes radiolabeling of short probes, and pp. 11.45-11.53, which describe hybridization conditions for oligonucleotide probes, including specific conditions for probe hybridization (pp. 11.50-11.51).

Methods of performing primer-directed amplification are also well-known in the art. Methods for performing the polymerase chain reaction (PCR) are compiled, *inter alia*, in McPherson, PCR Basics: From Background to Bench, Springer Verlag (2000); Innis *et al.* (eds.), PCR Applications: Protocols for Functional Genomics, Academic Press (1999); Gelfand *et al.* (eds.), PCR Strategies, Academic Press (1998); Newton *et al.*, PCR, Springer-Verlag New York (1997); Burke (ed.), PCR: Essential Techniques, John Wiley & Son Ltd (1996); White (ed.), PCR Cloning Protocols: From Molecular Cloning to Genetic Engineering, Vol. 67, Humana Press (1996); McPherson *et al.* (eds.), PCR 2: A Practical Approach, Oxford University Press, Inc. (1995); the disclosures of which are incorporated herein by reference in their entireties. Methods for performing RT-PCR are collected, *e.g.*, in Siebert *et al.* (eds.), Gene Cloning and Analysis by RT-PCR, Eaton Publishing Company/Bio Techniques Books Division, 1998; Siebert (ed.), PCR Technique:RT-PCR, Eaton Publishing Company/ BioTechniques Books (1995); the disclosure of which is incorporated herein by reference in its entirety.

PCR and hybridization methods may be used to identify and/or isolate allelic variants, homologous nucleic acid molecules and fragments of the nucleic acid molecules of the invention. PCR and hybridization methods may also be used to identify, amplify and/or isolate nucleic acid molecules that encode homologous proteins, analogs, fusion protein or muteins of the invention. The nucleic acid primers of the present invention can be used to prime amplification of nucleic acid molecules of the invention, using transcript-derived or genomic DNA as template.

The nucleic acid primers of the present invention can also be used, for example, to prime single base extension (SBE) for SNP detection (*See, e.g.*, U.S. Patent 6,004,744, the disclosure of which is incorporated herein by reference in its entirety).

Isothermal amplification approaches, such as rolling circle amplification, are also now well-described. *See, e.g.*, Schweitzer *et al.*, *Curr. Opin. Biotechnol.* 12(1): 21-7 (2001); U.S. Patents 5,854,033 and 5,714,320; and international patent publications WO 97/19193 and WO 00/15779, the disclosures of which are incorporated herein by

reference in their entireties. Rolling circle amplification can be combined with other techniques to facilitate SNP detection. *See, e.g.*, Lizardi *et al.*, *Nature Genet.* 19(3): 225-32 (1998).

Nucleic acid molecules of the present invention may be bound to a substrate either covalently or noncovalently. The substrate can be porous or solid, planar or non-planar, unitary or distributed. The bound nucleic acid molecules may be used as hybridization probes, and may be labeled or unlabeled. In a preferred embodiment, the bound nucleic acid molecules are unlabeled.

In one embodiment, the nucleic acid molecule of the present invention is bound to a porous substrate, *e.g.*, a membrane, typically comprising nitrocellulose, nylon, or positively-charged derivatized nylon. The nucleic acid molecule of the present invention can be used to detect a hybridizing nucleic acid molecule that is present within a labeled nucleic acid sample, *e.g.*, a sample of transcript-derived nucleic acids. In another embodiment, the nucleic acid molecule is bound to a solid substrate, including, without limitation, glass, amorphous silicon, crystalline silicon or plastics. Examples of plastics include, without limitation, polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof. The solid substrate may be any shape, including rectangular, disk-like and spherical. In a preferred embodiment, the solid substrate is a microscope slide or slide-shaped substrate.

The nucleic acid molecule of the present invention can be attached covalently to a surface of the support substrate or applied to a derivatized surface in a chaotropic agent that facilitates denaturation and adherence by presumed noncovalent interactions, or some combination thereof. The nucleic acid molecule of the present invention can be bound to a substrate to which a plurality of other nucleic acids are concurrently bound, hybridization to each of the plurality of bound nucleic acids being separately detectable. At low density, *e.g.* on a porous membrane, these substrate-bound collections are typically denominated macroarrays; at higher density, typically on a solid support, such as glass, these substrate bound collections of plural nucleic acids are colloquially termed microarrays. As used herein, the term microarray includes arrays of all densities. It is, therefore, another aspect of the invention to provide microarrays that include the nucleic acids of the present invention.

Expression Vectors, Host Cells and Recombinant Methods of Producing Polypeptides

Another aspect of the present invention relates to vectors that comprise one or more of the isolated nucleic acid molecules of the present invention, and host cells in which such vectors have been introduced.

The vectors can be used, *inter alia*, for propagating the nucleic acids of the present invention in host cells (cloning vectors), for shuttling the nucleic acids of the present invention between host cells derived from disparate organisms (shuttle vectors), for inserting the nucleic acids of the present invention into host cell chromosomes (insertion vectors), for expressing sense or antisense RNA transcripts of the nucleic acids of the present invention *in vitro* or within a host cell, and for expressing polypeptides encoded by the nucleic acids of the present invention, alone or as fusions to heterologous polypeptides (expression vectors). Vectors of the present invention will often be suitable for several such uses.

Vectors are by now well-known in the art, and are described, *inter alia*, in Jones *et al.* (eds.), <u>Vectors: Cloning Applications: Essential Techniques</u> (Essential Techniques Series), John Wiley & Son Ltd. (1998); Jones *et al.* (eds.), <u>Vectors: Expression Systems: Essential Techniques</u> (Essential Techniques Series), John Wiley & Son Ltd. (1998); Gacesa *et al.*, <u>Vectors: Essential Data</u>, John Wiley & Sons Ltd. (1995); Cid-Arregui (eds.), <u>Viral Vectors: Basic Science and Gene Therapy</u>, Eaton Publishing Co. (2000); Sambrook (2001), *supra*; Ausubel (1999), *supra*; the disclosures of which are incorporated herein by reference in their entireties. Furthermore, an enormous variety of vectors are available commercially. Use of existing vectors and modifications thereof being well within the skill in the art, only basic features need be described here.

Nucleic acid sequences may be expressed by operatively linking them to an expression control sequence in an appropriate expression vector and employing that expression vector to transform an appropriate unicellular host. Expression control sequences are sequences which control the transcription, post-transcriptional events and translation of nucleic acid sequences. Such operative linking of a nucleic sequence of this invention to an expression control sequence, of course, includes, if not already part of the nucleic acid sequence, the provision of a translation initiation codon, ATG or GTG, in the correct reading frame upstream of the nucleic acid sequence.

A wide variety of host/expression vector combinations may be employed in expressing the nucleic acid sequences of this invention. Useful expression vectors, for

example, may consist of segments of chromosomal, non-chromosomal and synthetic nucleic acid sequences.

In one embodiment, prokaryotic cells may be used with an appropriate vector. Prokaryotic host cells are often used for cloning and expression. In a preferred embodiment, prokaryotic host cells include *E. coli*, *Pseudomonas*, *Bacillus* and *Streptomyces*. In a preferred embodiment, bacterial host cells are used to express the nucleic acid molecules of the instant invention. Useful expression vectors for bacterial hosts include bacterial plasmids, such as those from *E. coli*, *Bacillus* or *Streptomyces*, including pBluescript, pGEX-2T, pUC vectors, col E1, pCR1, pBR322, pMB9 and their derivatives, wider host range plasmids, such as RP4, phage DNAs, *e.g.*, the numerous derivatives of phage lambda, *e.g.*, NM989, λGT10 and λGT11, and other phages, *e.g.*, M13 and filamentous single-stranded phage DNA. Where *E. coli* is used as host, selectable markers are, analogously, chosen for selectivity in gram negative bacteria: *e.g.*, typical markers confer resistance to antibiotics, such as ampicillin, tetracycline, chloramphenicol, kanamycin, streptomycin and zeocin; auxotrophic markers can also be used.

In other embodiments, eukaryotic host cells, such as yeast, insect, mammalian or plant cells, may be used. Yeast cells, typically *S. cerevisiae*, are useful for eukaryotic genetic studies, due to the ease of targeting genetic changes by homologous recombination and the ability to easily complement genetic defects using recombinantly expressed proteins. Yeast cells are useful for identifying interacting protein components, *e.g.* through use of a two-hybrid system. In a preferred embodiment, yeast cells are useful for protein expression. Vectors of the present invention for use in yeast will typically, but not invariably, contain an origin of replication suitable for use in yeast and a selectable marker that is functional in yeast. Yeast vectors include Yeast Integrating plasmids (*e.g.*, YIp5) and Yeast Replicating plasmids (the YRp and YEp series plasmids), Yeast Centromere plasmids (the YCp series plasmids), Yeast Artificial Chromosomes (YACs) which are based on yeast linear plasmids, denoted YLp, pGPD-2, 2μ plasmids and derivatives thereof, and improved shuttle vectors such as those described in Gietz *et al.*, *Gene*, 74: 527-34 (1988) (YIplac, YEplac and YCplac). Selectable markers in yeast vectors include a variety of auxotrophic markers, the most common of which are (in *Saccharomyces cerevisiae*) URA3, HIS3, LEU2, TRP1 and LYS2, which complement specific auxotrophic mutations, such as ura3-52, his3-D1, leu2-D1, trp1-D1 and lys2-201.

Insect cells are often chosen for high efficiency protein expression. Where the host cells are from *Spodoptera frugiperda* , *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA)), the vector replicative strategy is typically based upon the baculovirus life cycle. Typically, baculovirus transfer vectors are used to replace the wild-type AcMNPV polyhedrin gene with a heterologous gene of interest. Sequences that flank the polyhedrin gene in the wild-type genome are positioned 5' and 3' of the expression cassette on the transfer vectors. Following co-transfection with AcMNPV DNA, a homologous recombination event occurs between these sequences resulting in a recombinant virus carrying the gene of interest and the polyhedrin or p10 promoter. Selection can be based upon visual screening for lacZ fusion activity.

In another embodiment, the host cells may be mammalian cells, which are particularly useful for expression of proteins intended as pharmaceutical agents, and for screening of potential agonists and antagonists of a protein or a physiological pathway. Mammalian vectors intended for autonomous extrachromosomal replication will typically include a viral origin, such as the SV40 origin (for replication in cell lines expressing the large T-antigen, such as COS1 and COS7 cells), the papillomavirus origin, or the EBV origin for long term episomal replication (for use, *e.g.*, in 293-EBNA cells, which constitutively express the EBV EBNA-1 gene product and adenovirus E1A). Vectors intended for integration, and thus replication as part of the mammalian chromosome, can, but need not, include an origin of replication functional in mammalian cells, such as the SV40 origin. Vectors based upon viruses, such as adenovirus, adeno-associated virus, vaccinia virus, and various mammalian retroviruses, will typically replicate according to the viral replicative strategy. Selectable markers for use in mammalian cells include resistance to neomycin (G418), blasticidin, hygromycin and to zeocin, and selection based upon the purine salvage pathway using HAT medium.

Expression in mammalian cells can be achieved using a variety of plasmids, including pSV2, pBC12BI, and p91023, as well as lytic virus vectors (*e.g.*, vaccinia virus, adeno virus, and baculovirus), episomal virus vectors (*e.g.*, bovine papillomavirus), and retroviral vectors (*e.g.*, murine retroviruses). Useful vectors for insect cells include baculoviral vectors and pVL 941.

Plant cells can also be used for expression, with the vector replicon typically derived from a plant virus (*e.g.*, cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) and selectable markers chosen for suitability in plants.

It is known that codon usage of different host cells may be different. For example, a plant cell and a human cell may exhibit a difference in codon preference for encoding a particular amino acid. As a result, human mRNA may not be efficiently translated in a plant, bacteria or insect host cell. Therefore, another embodiment of this invention is directed to codon optimization. The codons of the nucleic acid molecules of the invention may be modified to resemble, as much as possible, genes naturally contained within the host cell without altering the amino acid sequence encoded by the nucleic acid molecule.

Any of a wide variety of expression control sequences may be used in these vectors to express the DNA sequences of this invention. Such useful expression control sequences include the expression control sequences associated with structural genes of the foregoing expression vectors. Expression control sequences that control transcription include, *e.g.*, promoters, enhancers and transcription termination sites. Expression control sequences in eukaryotic cells that control post-transcriptional events include splice donor and acceptor sites and sequences that modify the half-life of the transcribed RNA, *e.g.*, sequences that direct poly(A) addition or binding sites for RNA-binding proteins. Expression control sequences that control translation include ribosome binding sites, sequences which direct targeted expression of the polypeptide to or within particular cellular compartments, and sequences in the 5' and 3' untranslated regions that modify the rate or efficiency of translation.

Examples of useful expression control sequences for a prokaryote, *e.g.*, *E. coli*, will include a promoter, often a phage promoter, such as phage lambda pL promoter, the trc promoter, a hybrid derived from the trp and lac promoters, the bacteriophage T7 promoter (in *E. coli* cells engineered to express the T7 polymerase), the <u>TAC</u> or <u>TRC</u> system, the major operator and promoter regions of phage lambda, the control regions of fd coat protein, or the araBAD operon. Prokaryotic expression vectors may further include transcription terminators, such as the aspA terminator, and elements that facilitate translation, such as a consensus ribosome binding site and translation termination codon, Schomer *et al.*, *Proc. Natl. Acad. Sci. USA* 83: 8506-8510 (1986).

Expression control sequences for yeast cells, typically *S. cerevisiae*, will include a yeast promoter, such as the CYC1 promoter, the GAL1 promoter, the GAL10 promoter, ADH1 promoter, the promoters of the yeast $\underline{\alpha}$-mating system, or the GPD promoter, and will typically have elements that facilitate transcription termination, such as the transcription termination signals from the CYC1 or ADH1 gene.

Expression vectors useful for expressing proteins in mammalian cells will include a promoter active in mammalian cells. These promoters include those derived from mammalian viruses, such as the enhancer-promoter sequences from the immediate early gene of the human cytomegalovirus (CMV), the enhancer-promoter sequences from the Rous sarcoma virus long terminal repeat (RSV LTR), the enhancer-promoter from SV40 or the early and late promoters of adenovirus. Other expression control sequences include the promoter for 3-phosphoglycerate kinase or other glycolytic enzymes, the promoters of acid phosphatase. Other expression control sequences include those from the gene comprising the BSNA of interest. Often, expression is enhanced by incorporation of polyadenylation sites, such as the late SV40 polyadenylation site and the polyadenylation signal and transcription termination sequences from the bovine growth hormone (BGH) gene, and ribosome binding sites. Furthermore, vectors can include introns, such as intron II of rabbit β-globin gene and the SV40 splice elements.

Preferred nucleic acid vectors also include a selectable or amplifiable marker gene and means for amplifying the copy number of the gene of interest. Such marker genes are well-known in the art. Nucleic acid vectors may also comprise stabilizing sequences (*e.g.*, ori- or ARS-like sequences and telomere-like sequences), or may alternatively be designed to favor directed or non-directed integration into the host cell genome. In a preferred embodiment, nucleic acid sequences of this invention are inserted in frame into an expression vector that allows high level expression of an RNA which encodes a protein comprising the encoded nucleic acid sequence of interest. Nucleic acid cloning and sequencing methods are well-known to those of skill in the art and are described in an assortment of laboratory manuals, including Sambrook (1989), *supra*, Sambrook (2000), *supra*; and Ausubel (1992), *supra*, Ausubel (1999), *supra*. Product information from manufacturers of biological, chemical and immunological reagents also provide useful information.

Expression vectors may be either constitutive or inducible. Inducible vectors include either naturally inducible promoters, such as the trc promoter, which is regulated by the lac operon, and the pL promoter, which is regulated by tryptophan, the MMTV-LTR promoter, which is inducible by dexamethasone, or can contain synthetic promoters and/or additional elements that confer inducible control on adjacent promoters. Examples of inducible synthetic promoters are the hybrid Plac/ara-1 promoter and the PLtetO-1 promoter. The PltetO-1 promoter takes advantage of the high expression levels from the PL promoter of phage lambda, but replaces the lambda repressor sites with two

copies of operator 2 of the Tn10 tetracycline resistance operon, causing this promoter to be tightly repressed by the Tet repressor protein and induced in response to tetracycline (Tc) and Tc derivatives such as anhydrotetracycline. Vectors may also be inducible because they contain hormone response elements, such as the glucocorticoid response element (GRE) and the estrogen response element (ERE), which can confer hormone inducibility where vectors are used for expression in cells having the respective hormone receptors. To reduce background levels of expression, elements responsive to ecdysone, an insect hormone, can be used instead, with coexpression of the ecdysone receptor.

In one aspect of the invention, expression vectors can be designed to fuse the expressed polypeptide to small protein tags that facilitate purification and/or visualization. Tags that facilitate purification include a polyhistidine tag that facilitates purification of the fusion protein by immobilized metal affinity chromatography, for example using NiNTA resin (Qiagen Inc., Valencia, CA, USA) or TALON™ resin (cobalt immobilized affinity chromatography medium, Clontech Labs, Palo Alto, CA, USA). The fusion protein can include a chitin-binding tag and self-excising intein, permitting chitin-based purification with self-removal of the fused tag (IMPACT™ system, New England Biolabs, Inc., Beverley, MA, USA). Alternatively, the fusion protein can include a calmodulin-binding peptide tag, permitting purification by calmodulin affinity resin (Stratagene, La Jolla, CA, USA), or a specifically excisable fragment of the biotin carboxylase carrier protein, permitting purification of *in vivo* biotinylated protein using an avidin resin and subsequent tag removal (Promega, Madison, WI, USA). As another useful alternative, the proteins of the present invention can be expressed as a fusion protein with glutathione-S-transferase, the affinity and specificity of binding to glutathione permitting purification using glutathione affinity resins, such as Glutathione-Superflow Resin (Clontech Laboratories, Palo Alto, CA, USA), with subsequent elution with free glutathione. Other tags include, for example, the Xpress epitope, detectable by anti-Xpress antibody (Invitrogen, Carlsbad, CA, USA), a myc tag, detectable by anti-myc tag antibody, the V5 epitope, detectable by anti-V5 antibody (Invitrogen, Carlsbad, CA, USA), FLAG® epitope, detectable by anti-FLAG® antibody (Stratagene, La Jolla, CA, USA), and the HA epitope.

For secretion of expressed proteins, vectors can include appropriate sequences that encode secretion signals, such as leader peptides. For example, the pSecTag2 vectors (Invitrogen, Carlsbad, CA, USA) are 5.2 kb mammalian expression vectors that

carry the secretion signal from the V-J2-C region of the mouse Ig kappa-chain for efficient secretion of recombinant proteins from a variety of mammalian cell lines.

Expression vectors can also be designed to fuse proteins encoded by the heterologous nucleic acid insert to polypeptides that are larger than purification and/or identification tags. Useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusion to intrinsically fluorescent proteins, such as those that have a green fluorescent protein (GFP)-like chromophore, fusions to the IgG Fc region, and fusion proteins for use in two hybrid systems.

Vectors for phage display fuse the encoded polypeptide to, *e.g.*, the gene III protein (pIII) or gene VIII protein (pVIII) for display on the surface of filamentous phage, such as M13. *See* Barbas *et al.*, Phage Display: A Laboratory Manual, Cold Spring Harbor Laboratory Press (2001); Kay *et al.* (eds.), Phage Display of Peptides and Proteins: A Laboratory Manual, Academic Press, Inc., (1996); Abelson *et al.* (eds.), Combinatorial Chemistry (Methods in Enzymology, Vol. 267) Academic Press (1996). Vectors for yeast display, *e.g.* the pYD1 yeast display vector (Invitrogen, Carlsbad, CA, USA), use the α-agglutinin yeast adhesion receptor to display recombinant protein on the surface of *S. cerevisiae*. Vectors for mammalian display, *e.g.*, the pDisplay™ vector (Invitrogen, Carlsbad, CA, USA), target recombinant proteins using an N-terminal cell surface targeting signal and a C-terminal transmembrane anchoring domain of platelet derived growth factor receptor.

A wide variety of vectors now exist that fuse proteins encoded by heterologous nucleic acids to the chromophore of the substrate-independent, intrinsically fluorescent green fluorescent protein from *Aequorea victoria* ("GFP") and its variants. The GFP-like chromophore can be selected from GFP-like chromophores found in naturally occurring proteins, such as *A. victoria* GFP (GenBank accession number AAA27721), *Renilla reniformis* GFP, FP583 (GenBank accession no. AF168419) (DsRed), FP593 (AF272711), FP483 (AF168420), FP484 (AF168424), FP595 (AF246709), FP486 (AF168421), FP538 (AF168423), and FP506 (AF168422), and need include only so much of the native protein as is needed to retain the chromophore's intrinsic fluorescence. Methods for determining the minimal domain required for fluorescence are known in the art. *See* Li *et al.*, *J. Biol. Chem.* 272: 28545-28549 (1997). Alternatively, the GFP-like chromophore can be selected from GFP-like chromophores modified from those found in nature. The methods for engineering such modified GFP-like chromophores and testing them for fluorescence activity, both alone and as part of

protein fusions, are well-known in the art. *See* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Palm *et al.*, *Methods Enzymol.* 302: 378-394 (1999), incorporated herein by reference in its entirety. A variety of such modified chromophores are now commercially available and can readily be used in the fusion proteins of the present invention. These include EGFP ("enhanced GFP"), EBFP ("enhanced blue fluorescent protein"), BFP2, EYFP ("enhanced yellow fluorescent protein"), ECFP ("enhanced cyan fluorescent protein") or Citrine. EGFP (*see, e.g,* Cormack *et al.*, *Gene* 173: 33–38 (1996); United States Patent Nos. 6,090,919 and 5,804,387) is found on a variety of vectors, both plasmid and viral, which are available commercially (Clontech Labs, Palo Alto, CA, USA); EBFP is optimized for expression in mammalian cells whereas BFP2, which retains the original jellyfish codons, can be expressed in bacteria (*see, e.g,.* Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996) and Cormack *et al.*, *Gene* 173: 33-38 (1996)). Vectors containing these blue-shifted variants are available from Clontech Labs (Palo Alto, CA, USA). Vectors containing EYFP, ECFP (*see, e.g.*, Heim *et al.*, *Curr. Biol.* 6: 178-182 (1996); Miyawaki *et al.*, *Nature* 388: 882-887 (1997)) and Citrine (*see, e.g.*, Heikal *et al.*, *Proc. Natl. Acad. Sci. USA* 97: 11996-12001 (2000)) are also available from Clontech Labs. The GFP-like chromophore can also be drawn from other modified GFPs, including those described in U.S. Patents 6,124,128; 6,096,865; 6,090,919; 6,066,476; 6,054,321; 6,027,881; 5,968,750; 5,874,304; 5,804,387; 5,777,079; 5,741,668; and 5,625,048, the disclosures of which are incorporated herein by reference in their entireties. *See* also Conn (ed.), Green Fluorescent Protein (Methods in Enzymology, Vol. 302), Academic Press, Inc. (1999). The GFP-like chromophore of each of these GFP variants can usefully be included in the fusion proteins of the present invention.

Fusions to the IgG Fc region increase serum half life of protein pharmaceutical products through interaction with the FcRn receptor (also denominated the FcRp receptor and the Brambell receptor, FcRb), further described in International Patent Application Nos. WO 97/43316, WO 97/34631, WO 96/32478, WO 96/18412.

For long-term, high-yield recombinant production of the proteins, protein fusions, and protein fragments of the present invention, stable expression is preferred. Stable expression is readily achieved by integration into the host cell genome of vectors having selectable markers, followed by selection of these integrants. Vectors such as pUB6/V5-His A, B, and C (Invitrogen, Carlsbad, CA, USA) are designed for high-level stable expression of heterologous proteins in a wide range of mammalian tissue types and

cell lines. pUB6/V5-His uses the promoter/enhancer sequence from the human ubiquitin C gene to drive expression of recombinant proteins: expression levels in 293, CHO, and NIH3T3 cells are comparable to levels from the CMV and human EF-1a promoters. The bsd gene permits rapid selection of stably transfected mammalian cells with the potent antibiotic blasticidin.

Replication incompetent retroviral vectors, typically derived from Moloney murine leukemia virus, also are useful for creating stable transfectants having integrated provirus. The highly efficient transduction machinery of retroviruses, coupled with the availability of a variety of packaging cell lines such as RetroPack™ PT 67, EcoPack2™-293, AmphoPack-293, and GP2-293 cell lines (all available from Clontech Laboratories, Palo Alto, CA, USA), allow a wide host range to be infected with high efficiency; varying the multiplicity of infection readily adjusts the copy number of the integrated provirus.

Of course, not all vectors and expression control sequences will function equally well to express the nucleic acid sequences of this invention. Neither will all hosts function equally well with the same expression system. However, one of skill in the art may make a selection among these vectors, expression control sequences and hosts without undue experimentation and without departing from the scope of this invention. For example, in selecting a vector, the host must be considered because the vector must be replicated in it. The vector's copy number, the ability to control that copy number, the ability to control integration, if any, and the expression of any other proteins encoded by the vector, such as antibiotic or other selection markers, should also be considered. The present invention further includes host cells comprising the vectors of the present invention, either present episomally within the cell or integrated, in whole or in part, into the host cell chromosome. Among other considerations, some of which are described above, a host cell strain may be chosen for its ability to process the expressed protein in the desired fashion. Such post-translational modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation, and it is an aspect of the present invention to provide BSPs with such post-translational modifications.

Polypeptides of the invention may be post-translationally modified. Post-translational modifications include phosphorylation of amino acid residues serine, threonine and/or tyrosine, N-linked and/or O-linked glycosylation, methylation, acetylation, prenylation, methylation, acetylation, arginylation, ubiquination and

racemization. One may determine whether a polypeptide of the invention is likely to be post-translationally modified by analyzing the sequence of the polypeptide to determine if there are peptide motifs indicative of sites for post-translational modification. There are a number of computer programs that permit prediction of post-translational modifications. See, e.g., www.expasy.org (accessed August 31, 2001), which includes PSORT, for prediction of protein sorting signals and localization sites, SignalP, for prediction of signal peptide cleavage sites, MITOPROT and Predotar, for prediction of mitochondrial targeting sequences, NetOGlyc, for prediction of type O-glycosylation sites in mammalian proteins, big-PI Predictor and DGPI, for prediction of prenylation-anchor and cleavage sites, and NetPhos, for prediction of Ser, Thr and Tyr phosphorylation sites in eukaryotic proteins. Other computer programs, such as those included in GCG, also may be used to determine post-translational modification peptide motifs.

General examples of types of post-translational modifications may be found in web sites such as the Delta Mass database http://www.abrf.org/ABRF/Research Committees/deltamass/deltamass.html (accessed October 19, 2001); "GlycoSuiteDB: a new curated relational database of glycoprotein glycan structures and their biological sources" Cooper et al. Nucleic Acids Res. 29; 332-335 (2001) and http://www.glycosuite.com/ (accessed October 19, 2001); "O-GLYCBASE version 4.0: a revised database of O-glycosylated proteins" Gupta et al. Nucleic Acids Research, 27: 370-372 (1999) and http://www.cbs.dtu.dk/databases/OGLYCBASE/ (accessed October 19, 2001); "PhosphoBase, a database of phosphorylation sites: release 2.0.", Kreegipuu et al. Nucleic Acids Res 27(1):237-239 (1999) and http://www.cbs.dtu.dk/ databases/PhosphoBase/ (accessed October 19, 2001); or http://pir.georgetown.edu/ pirwww/search/textresid.html (accessed October 19, 2001).

Tumorigenesis is often accompanied by alterations in the post-translational modifications of proteins. Thus, in another embodiment, the invention provides polypeptides from cancerous cells or tissues that have altered post-translational modifications compared to the post-translational modifications of polypeptides from normal cells or tissues. A number of altered post-translational modifications are known. One common alteration is a change in phosphorylation state, wherein the polypeptide from the cancerous cell or tissue is hyperphosphorylated or hypophosphorylated compared to the polypeptide from a normal tissue, or wherein the polypeptide is phosphorylated on different residues than the polypeptide from a normal cell. Another

common alteration is a change in glycosylation state, wherein the polypeptide from the cancerous cell or tissue has more or less glycosylation than the polypeptide from a normal tissue, and/or wherein the polypeptide from the cancerous cell or tissue has a different type of glycosylation than the polypeptide from a noncancerous cell or tissue. Changes in glycosylation may be critical because carbohydrate-protein and carbohydrate-carbohydrate interactions are important in cancer cell progression, dissemination and invasion. See, e.g., Barchi, *Curr. Pharm. Des.* 6: 485-501 (2000), Verma, *Cancer Biochem. Biophys.* 14: 151-162 (1994) and Dennis et al., *Bioessays* 5: 412-421 (1999).

Another post-translational modification that may be altered in cancer cells is prenylation. Prenylation is the covalent attachment of a hydrophobic prenyl group (either farnesyl or geranylgeranyl) to a polypeptide. Prenylation is required for localizing a protein to a cell membrane and is often required for polypeptide function. For instance, the Ras superfamily of GTPase signaling proteins must be prenylated for function in a cell. See, e.g., Prendergast et al., *Semin. Cancer Biol.* 10: 443-452 (2000) and Khwaja et al., *Lancet* 355: 741-744 (2000).

Other post-translation modifications that may be altered in cancer cells include, without limitation, polypeptide methylation, acetylation, arginylation or racemization of amino acid residues. In these cases, the polypeptide from the cancerous cell may exhibit either increased or decreased amounts of the post-translational modification compared to the corresponding polypeptides from noncancerous cells.

Other polypeptide alterations in cancer cells include abnormal polypeptide cleavage of proteins and aberrant protein-protein interactions. Abnormal polypeptide cleavage may be cleavage of a polypeptide in a cancerous cell that does not usually occur in a normal cell, or a lack of cleavage in a cancerous cell, wherein the polypeptide is cleaved in a normal cell. Aberrant protein-protein interactions may be either covalent cross-linking or non-covalent binding between proteins that do not normally bind to each other. Alternatively, in a cancerous cell, a protein may fail to bind to another protein to which it is bound in a noncancerous cell. Alterations in cleavage or in protein-protein interactions may be due to over- or underproduction of a polypeptide in a cancerous cell compared to that in a normal cell, or may be due to alterations in post-translational modifications (see above) of one or more proteins in the cancerous cell. See, e.g., Henschen-Edman, *Ann. N.Y. Acad. Sci.* 936: 580-593 (2001).

Alterations in polypeptide post-translational modifications, as well as changes in polypeptide cleavage and protein-protein interactions, may be determined by any method

known in the art. For instance, alterations in phosphorylation may be determined by using anti-phosphoserine, anti-phosphothreonine or anti-phosphotyrosine antibodies or by amino acid analysis. Glycosylation alterations may be determined using antibodies specific for different sugar residues, by carbohydrate sequencing, or by alterations in the size of the glycoprotein, which can be determined by, e.g., SDS polyacrylamide gel electrophoresis (PAGE). Other alterations of post-translational modifications, such as prenylation, racemization, methylation, acetylation and arginylation, may be determined by chemical analysis, protein sequencing, amino acid analysis, or by using antibodies specific for the particular post-translational modifications. Changes in protein-protein interactions and in polypeptide cleavage may be analyzed by any method known in the art including, without limitation, non-denaturing PAGE (for non-covalent protein-protein interactions), SDS PAGE (for covalent protein-protein interactions and protein cleavage), chemical cleavage, protein sequencing or immunoassays.

In another embodiment, the invention provides polypeptides that have been post-translationally modified. In one embodiment, polypeptides may be modified enzymatically or chemically, by addition or removal of a post-translational modification. For example, a polypeptide may be glycosylated or deglycosylated enzymatically. Similarly, polypeptides may be phosphorylated using a purified kinase, such as a MAP kinase (e.g, p38, ERK, or JNK) or a tyrosine kinase (e.g., Src or erbB2). A polypeptide may also be modified through synthetic chemistry. Alternatively, one may isolate the polypeptide of interest from a cell or tissue that expresses the polypeptide with the desired post-translational modification. In another embodiment, a nucleic acid molecule encoding the polypeptide of interest is introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide in the desired fashion. If the polypeptide does not contain a motif for a desired post-translational modification, one may alter the post-translational modification by mutating the nucleic acid sequence of a nucleic acid molecule encoding the polypeptide so that it contains a site for the desired post-translational modification. Amino acid sequences that may be post-translationally modified are known in the art. See, e.g., the programs described above on the website www.expasy.org. The nucleic acid molecule is then be introduced into a host cell that is capable of post-translationally modifying the encoded polypeptide. Similarly, one may delete sites that are post-translationally modified by either mutating the nucleic acid sequence so that the encoded polypeptide does not contain the post-translational

modification motif, or by introducing the native nucleic acid molecule into a host cell that is not capable of post-translationally modifying the encoded polypeptide.

In selecting an expression control sequence, a variety of factors should also be considered. These include, for example, the relative strength of the sequence, its controllability, and its compatibility with the nucleic acid sequence of this invention, particularly with regard to potential secondary structures. Unicellular hosts should be selected by consideration of their compatibility with the chosen vector, the toxicity of the product coded for by the nucleic acid sequences of this invention, their secretion characteristics, their ability to fold the polypeptide correctly, their fermentation or culture requirements, and the ease of purification from them of the products coded for by the nucleic acid sequences of this invention.

The recombinant nucleic acid molecules and more particularly, the expression vectors of this invention may be used to express the polypeptides of this invention as recombinant polypeptides in a heterologous host cell. The polypeptides of this invention may be full-length or less than full-length polypeptide fragments recombinantly expressed from the nucleic acid sequences according to this invention. Such polypeptides include analogs, derivatives and muteins that may or may not have biological activity.

Vectors of the present invention will also often include elements that permit *in vitro* transcription of RNA from the inserted heterologous nucleic acid. Such vectors typically include a phage promoter, such as that from T7, T3, or SP6, flanking the nucleic acid insert. Often two different such promoters flank the inserted nucleic acid, permitting separate *in vitro* production of both sense and antisense strands.

Transformation and other methods of introducing nucleic acids into a host cell (*e.g.*, conjugation, protoplast transformation or fusion, transfection, electroporation, liposome delivery, membrane fusion techniques, high velocity DNA-coated pellets, viral infection and protoplast fusion) can be accomplished by a variety of methods which are well-known in the art (*See*, for instance, Ausubel, *supra*, and Sambrook *et al.*, *supra*). Bacterial, yeast, plant or mammalian cells are transformed or transfected with an expression vector, such as a plasmid, a cosmid, or the like, wherein the expression vector comprises the nucleic acid of interest. Alternatively, the cells may be infected by a viral expression vector comprising the nucleic acid of interest. Depending upon the host cell, vector, and method of transformation used, transient or stable expression of the polypeptide will be constitutive or inducible. One having ordinary skill in the art will be

able to decide whether to express a polypeptide transiently or stably, and whether to express the protein constitutively or inducibly.

A wide variety of unicellular host cells are useful in expressing the DNA sequences of this invention. These hosts may include well-known eukaryotic and prokaryotic hosts, such as strains of, fungi, yeast, insect cells such as *Spodoptera frugiperda* (SF9), animal cells such as CHO, as well as plant cells in tissue culture. Representative examples of appropriate host cells include, but are not limited to, bacterial cells, such as *E. coli*, *Caulobacter crescentus*, *Streptomyces* species, and *Salmonella typhimurium*; yeast cells, such as *Saccharomyces cerevisiae, Schizosaccharomyces pombe*, *Pichia pastoris*, *Pichia methanolica*; insect cell lines, such as those from *Spodoptera frugiperda*, *e.g.*, Sf9 and Sf21 cell lines, and expresSF™ cells (Protein Sciences Corp., Meriden, CT, USA), *Drosophila* S2 cells, and *Trichoplusia ni* High Five® Cells (Invitrogen, Carlsbad, CA, USA); and mammalian cells. Typical mammalian cells include BHK cells, BSC 1 cells, BSC 40 cells, BMT 10 cells, VERO cells, COS1 cells, COS7 cells, Chinese hamster ovary (CHO) cells, 3T3 cells, NIH 3T3 cells, 293 cells, HEPG2 cells, HeLa cells, L cells, MDCK cells, HEK293 cells, WI38 cells, murine ES cell lines (*e.g.*, from strains 129/SV, C57/BL6, DBA-1, 129/SVJ), K562 cells, Jurkat cells, and BW5147 cells. Other mammalian cell lines are well-known and readily available from the American Type Culture Collection (ATCC) (Manassas, VA, USA) and the National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Cell Repositories (Camden, NJ, USA). Cells or cell lines derived from breast are particularly preferred because they may provide a more native post-translational processing. Particularly preferred are human breast cells.

Particular details of the transfection, expression and purification of recombinant proteins are well documented and are understood by those of skill in the art. Further details on the various technical aspects of each of the steps used in recombinant production of foreign genes in bacterial cell expression systems can be found in a number of texts and laboratory manuals in the art. *See, e.g.*, Ausubel (1992), *supra*, Ausubel (1999), *supra*, Sambrook (1989), *supra*, and Sambrook (2001), *supra*, herein incorporated by reference.

Methods for introducing the vectors and nucleic acids of the present invention into the host cells are well-known in the art; the choice of technique will depend primarily upon the specific vector to be introduced and the host cell chosen.

Nucleic acid molecules and vectors may be introduced into prokaryotes, such as *E. coli,* in a number of ways. For instance, phage lambda vectors will typically be packaged using a packaging extract (*e.g.*, Gigapack® packaging extract, Stratagene, La Jolla, CA, USA), and the packaged virus used to infect *E. coli.*

Plasmid vectors will typically be introduced into chemically competent or electrocompetent bacterial cells. *E. coli* cells can be rendered chemically competent by treatment, *e.g.*, with $CaCl_2$, or a solution of Mg^{2+}, Mn^{2+}, Ca^{2+}, Rb^+ or K^+, dimethyl sulfoxide, dithiothreitol, and hexamine cobalt (III), Hanahan, *J. Mol. Biol.* 166(4):557-80 (1983), and vectors introduced by heat shock. A wide variety of chemically competent strains are also available commercially (*e.g.*, Epicurian Coli® XL10-Gold® Ultracompetent Cells (Stratagene, La Jolla, CA, USA); DH5α competent cells (Clontech Laboratories, Palo Alto, CA, USA); and TOP10 Chemically Competent E. coli Kit (Invitrogen, Carlsbad, CA, USA)). Bacterial cells can be rendered electrocompetent, that is, competent to take up exogenous DNA by electroporation, by various pre-pulse treatments; vectors are introduced by electroporation followed by subsequent outgrowth in selected media. An extensive series of protocols is provided online in Electroprotocols (BioRad, Richmond, CA, USA) (http://www.biorad.com/LifeScience/pdf/New_Gene_Pulser.pdf).

Vectors can be introduced into yeast cells by spheroplasting, treatment with lithium salts, electroporation, or protoplast fusion. Spheroplasts are prepared by the action of hydrolytic enzymes such as snail-gut extract, usually denoted Glusulase, or Zymolyase, an enzyme from *Arthrobacter luteus*, to remove portions of the cell wall in the presence of osmotic stabilizers, typically 1 M sorbitol. DNA is added to the spheroplasts, and the mixture is co-precipitated with a solution of polyethylene glycol (PEG) and Ca^{2+}. Subsequently, the cells are resuspended in a solution of sorbitol, mixed with molten agar and then layered on the surface of a selective plate containing sorbitol.

For lithium-mediated transformation, yeast cells are treated with lithium acetate, which apparently permeabilizes the cell wall, DNA is added and the cells are co-precipitated with PEG. The cells are exposed to a brief heat shock, washed free of PEG and lithium acetate, and subsequently spread on plates containing ordinary selective medium. Increased frequencies of transformation are obtained by using specially-prepared single-stranded carrier DNA and certain organic solvents. Schiestl *et al.*, *Curr. Genet.* 16(5-6): 339-46 (1989).

For electroporation, freshly-grown yeast cultures are typically washed, suspended in an osmotic protectant, such as sorbitol, mixed with DNA, and the cell suspension pulsed in an electroporation device. Subsequently, the cells are spread on the surface of plates containing selective media. Becker *et al.*, *Methods Enzymol.* 194: 182-187 (1991). The efficiency of transformation by electroporation can be increased over 100-fold by using PEG, single-stranded carrier DNA and cells that are in late log-phase of growth. Larger constructs, such as YACs, can be introduced by protoplast fusion.

Mammalian and insect cells can be directly infected by packaged viral vectors, or transfected by chemical or electrical means. For chemical transfection, DNA can be coprecipitated with $CaPO_4$ or introduced using liposomal and nonliposomal lipid-based agents. Commercial kits are available for $CaPO_4$ transfection (CalPhos™ Mammalian Transfection Kit, Clontech Laboratories, Palo Alto, CA, USA), and lipid-mediated transfection can be practiced using commercial reagents, such as LIPOFECTAMINE™ 2000, LIPOFECTAMINE™ Reagent, CELLFECTIN® Reagent, and LIPOFECTIN® Reagent (Invitrogen, Carlsbad, CA, USA), DOTAP Liposomal Transfection Reagent, FuGENE 6, X-tremeGENE Q2, DOSPER, (Roche Molecular Biochemicals, Indianapolis, IN USA), Effectene™, PolyFect®, Superfect® (Qiagen, Inc., Valencia, CA, USA). Protocols for electroporating mammalian cells can be found online in Electroprotocols (Bio-Rad, Richmond, CA, USA) (http://www.bio-rad.com/LifeScience/pdf/ New_Gene_Pulser.pdf); Norton *et al.* (eds.), Gene Transfer Methods: Introducing DNA into Living Cells and Organisms, BioTechniques Books, Eaton Publishing Co. (2000); incorporated herein by reference in its entirety. Other transfection techniques include transfection by particle bombardment and microinjection. *See, e.g.*, Cheng *et al.*, *Proc. Natl. Acad. Sci. USA* 90(10): 4455-9 (1993); Yang *et al.*, *Proc. Natl. Acad. Sci. USA* 87(24): 9568-72 (1990).

Production of the recombinantly produced proteins of the present invention can optionally be followed by purification.

Purification of recombinantly expressed proteins is now well by those skilled in the art. *See*, *e.g.*, Thorner *et al.* (eds.), Applications of Chimeric Genes and Hybrid Proteins, Part A: Gene Expression and Protein Purification (Methods in Enzymology, Vol. 326), Academic Press (2000); Harbin (ed.), Cloning, Gene Expression and Protein Purification : Experimental Procedures and Process Rationale, Oxford Univ. Press (2001); Marshak *et al.*, Strategies for Protein Purification and Characterization: A Laboratory Course Manual, Cold Spring Harbor Laboratory Press (1996); and Roe (ed.),

Protein Purification Applications, Oxford University Press (2001); the disclosures of which are incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, if purification tags have been fused through use of an expression vector that appends such tags, purification can be effected, at least in part, by means appropriate to the tag, such as use of immobilized metal affinity chromatography for polyhistidine tags. Other techniques common in the art include ammonium sulfate fractionation, immunoprecipitation, fast protein liquid chromatography (FPLC), high performance liquid chromatography (HPLC), and preparative gel electrophoresis.

Polypeptides

Another object of the invention is to provide polypeptides encoded by the nucleic acid molecules of the instant invention. In a preferred embodiment, the polypeptide is a breast specific polypeptide (BSP). In an even more preferred embodiment, the polypeptide is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 172 through 295. A polypeptide as defined herein may be produced recombinantly, as discussed *supra*, may be isolated from a cell that naturally expresses the protein, or may be chemically synthesized following the teachings of the specification and using methods well-known to those having ordinary skill in the art.

In another aspect, the polypeptide may comprise a fragment of a polypeptide, wherein the fragment is as defined herein. In a preferred embodiment, the polypeptide fragment is a fragment of a BSP. In a more preferred embodiment, the fragment is derived from a polypeptide comprising the amino acid sequence of SEQ ID NO: 172 through 295. A polypeptide that comprises only a fragment of an entire BSP may or may not be a polypeptide that is also a BSP. For instance, a full-length polypeptide may be breast-specific, while a fragment thereof may be found in other tissues as well as in breast. A polypeptide that is not a BSP, whether it is a fragment, analog, mutein, homologous protein or derivative, is nevertheless useful, especially for immunizing animals to prepare anti-BSP antibodies. However, in a preferred embodiment, the part or fragment is a BSP. Methods of determining whether a polypeptide is a BSP are described *infra*.

Fragments of at least 6 contiguous amino acids are useful in mapping B cell and T cell epitopes of the reference protein. *See, e.g.*, Geysen *et al.*, *Proc. Natl. Acad. Sci. USA* 81: 3998-4002 (1984) and U.S. Patents 4,708,871 and 5,595,915, the disclosures of

which are incorporated herein by reference in their entireties. Because the fragment need not itself be immunogenic, part of an immunodominant epitope, nor even recognized by native antibody, to be useful in such epitope mapping, all fragments of at least 6 amino acids of the proteins of the present invention have utility in such a study.

Fragments of at least 8 contiguous amino acids, often at least 15 contiguous amino acids, are useful as immunogens for raising antibodies that recognize the proteins of the present invention. *See, e.g.*, Lerner, Nature 299: 592-596 (1982); Shinnick *et al.*, *Annu. Rev. Microbiol.* 37: 425-46 (1983); Sutcliffe *et al.*, *Science* 219: 660-6 (1983), the disclosures of which are incorporated herein by reference in their entireties. As further described in the above-cited references, virtually all 8-mers, conjugated to a carrier, such as a protein, prove immunogenic, meaning that they are capable of eliciting antibody for the conjugated peptide; accordingly, all fragments of at least 8 amino acids of the proteins of the present invention have utility as immunogens.

Fragments of at least 8, 9, 10 or 12 contiguous amino acids are also useful as competitive inhibitors of binding of the entire protein, or a portion thereof, to antibodies (as in epitope mapping), and to natural binding partners, such as subunits in a multimeric complex or to receptors or ligands of the subject protein; this competitive inhibition permits identification and separation of molecules that bind specifically to the protein of interest, U.S. Patents 5,539,084 and 5,783,674, incorporated herein by reference in their entireties.

The protein, or protein fragment, of the present invention is thus at least 6 amino acids in length, typically at least 8, 9, 10 or 12 amino acids in length, and often at least 15 amino acids in length. Often, the protein of the present invention, or fragment thereof, is at least 20 amino acids in length, even 25 amino acids, 30 amino acids, 35 amino acids, or 50 amino acids or more in length. Of course, larger fragments having at least 75 amino acids, 100 amino acids, or even 150 amino acids are also useful, and at times preferred.

One having ordinary skill in the art can produce fragments of a polypeptide by truncating the nucleic acid molecule, *e.g.*, a BSNA, encoding the polypeptide and then expressing it recombinantly. Alternatively, one can produce a fragment by chemically synthesizing a portion of the full-length polypeptide. One may also produce a fragment by enzymatically cleaving either a recombinant polypeptide or an isolated naturally-occurring polypeptide. Methods of producing polypeptide fragments are well-known in the art. *See, e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992),

supra; and Ausubel (1999), *supra*. In one embodiment, a polypeptide comprising only a fragment of polypeptide of the invention, preferably a BSP, may be produced by chemical or enzymatic cleavage of a polypeptide. In a preferred embodiment, a polypeptide fragment is produced by expressing a nucleic acid molecule encoding a fragment of the polypeptide, preferably a BSP, in a host cell.

By "polypeptides" as used herein it is also meant to be inclusive of mutants, fusion proteins, homologous proteins and allelic variants of the polypeptides specifically exemplified.

A mutant protein, or mutein, may have the same or different properties compared to a naturally-occurring polypeptide and comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of a native protein. Small deletions and insertions can often be found that do not alter the function of the protein. In one embodiment, the mutein may or may not be breast-specific. In a preferred embodiment, the mutein is breast-specific. In a preferred embodiment, the mutein is a polypeptide that comprises at least one amino acid insertion, duplication, deletion, rearrangement or substitution compared to the amino acid sequence of SEQ ID NO: 172 through 295. In a more preferred embodiment, the mutein is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 172 through 295. In yet a more preferred embodiment, the mutein exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97%, 98%, 99% or 99.5% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 172 through 295.

A mutein may be produced by isolation from a naturally-occurring mutant cell, tissue or organism. A mutein may be produced by isolation from a cell, tissue or organism that has been experimentally mutagenized. Alternatively, a mutein may be produced by chemical manipulation of a polypeptide, such as by altering the amino acid residue to another amino acid residue using synthetic or semi-synthetic chemical techniques. In a preferred embodiment, a mutein may be produced from a host cell comprising an altered nucleic acid molecule compared to the naturally-occurring nucleic acid molecule. For instance, one may produce a mutein of a polypeptide by introducing one or more mutations into a nucleic acid sequence of the invention and then expressing it recombinantly. These mutations may be targeted, in which particular encoded amino

acids are altered, or may be untargeted, in which random encoded amino acids within the polypeptide are altered. Muteins with random amino acid alterations can be screened for a particular biological activity or property, particularly whether the polypeptide is breast-specific, as described below. Multiple random mutations can be introduced into the gene by methods well-known to the art, *e.g.*, by error-prone PCR, shuffling, oligonucleotide-directed mutagenesis, assembly PCR, sexual PCR mutagenesis, *in vivo* mutagenesis, cassette mutagenesis, recursive ensemble mutagenesis, exponential ensemble mutagenesis and site-specific mutagenesis. Methods of producing muteins with targeted or random amino acid alterations are well-known in the art. *See, e.g.*, Sambrook (1989), *supra;* Sambrook (2001), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), U.S. Patent 5,223,408, and the references discussed *supra*, each herein incorporated by reference.

By "polypeptide" as used herein it is also meant to be inclusive of polypeptides homologous to those polypeptides exemplified herein. In a preferred embodiment, the polypeptide is homologous to a BSP. In an even more preferred embodiment, the polypeptide is homologous to a BSP selected from the group having an amino acid sequence of SEQ ID NO: 172 through 295. In a preferred embodiment, the homologous polypeptide is one that exhibits significant sequence identity to a BSP. In a more preferred embodiment, the polypeptide is one that exhibits significant sequence identity to an comprising an amino acid sequence of SEQ ID NO: 172 through 295. In an even more preferred embodiment, the homologous polypeptide is one that exhibits at least 50% sequence identity, more preferably at least 60% sequence identity, even more preferably at least 70%, yet more preferably at least 80% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 172 through 295. In a yet more preferred embodiment, the homologous polypeptide is one that exhibits at least 85%, more preferably 90%, even more preferably 95% or 96%, and yet more preferably at least 97% or 98% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 172 through 295. In another preferred embodiment, the homologous polypeptide is one that exhibits at least 99%, more preferably 99.5%, even more preferably 99.6%, 99.7%, 99.8% or 99.9% sequence identity to a BSP comprising an amino acid sequence of SEQ ID NO: 172 through 295. In a preferred embodiment, the amino acid substitutions are conservative amino acid substitutions as discussed above.

In another embodiment, the homologous polypeptide is one that is encoded by a nucleic acid molecule that selectively hybridizes to a BSNA. In a preferred embodiment,

the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a BSNA under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the BSNA is selected from the group consisting of SEQ ID NO: 1 through 171. In another preferred embodiment, the homologous polypeptide is encoded by a nucleic acid molecule that hybridizes to a nucleic acid molecule that encodes a BSP under low stringency, moderate stringency or high stringency conditions, as defined herein. In a more preferred embodiment, the BSP is selected from the group consisting of SEQ ID NO: 172 through 295.

The homologous polypeptide may be a naturally-occurring one that is derived from another species, especially one derived from another primate, such as chimpanzee, gorilla, rhesus macaque, baboon or gorilla, wherein the homologous polypeptide comprises an amino acid sequence that exhibits significant sequence identity to that of SEQ ID NO: 172 through 295. The homologous polypeptide may also be a naturally-occurring polypeptide from a human, when the BSP is a member of a family of polypeptides. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-primate, mammalian species, including without limitation, domesticated species, *e.g.*, dog, cat, mouse, rat, rabbit, guinea pig, hamster, cow, horse, goat or pig. The homologous polypeptide may also be a naturally-occurring polypeptide derived from a non-mammalian species, such as birds or reptiles. The naturally-occurring homologous protein may be isolated directly from humans or other species. Alternatively, the nucleic acid molecule encoding the naturally-occurring homologous polypeptide may be isolated and used to express the homologous polypeptide recombinantly. In another embodiment, the homologous polypeptide may be one that is experimentally produced by random mutation of a nucleic acid molecule and subsequent expression of the nucleic acid molecule. In another embodiment, the homologous polypeptide may be one that is experimentally produced by directed mutation of one or more codons to alter the encoded amino acid of a BSP. Further, the homologous protein may or may not encode polypeptide that is a BSP. However, in a preferred embodiment, the homologous polypeptide encodes a polypeptide that is a BSP.

Relatedness of proteins can also be characterized using a second functional test, the ability of a first protein competitively to inhibit the binding of a second protein to an antibody. It is, therefore, another aspect of the present invention to provide isolated proteins not only identical in sequence to those described with particularity herein, but also to provide isolated proteins ("cross-reactive proteins") that competitively inhibit the

binding of antibodies to all or to a portion of various of the isolated polypeptides of the present invention. Such competitive inhibition can readily be determined using immunoassays well-known in the art.

As discussed above, single nucleotide polymorphisms (SNPs) occur frequently in eukaryotic genomes, and the sequence determined from one individual of a species may differ from other allelic forms present within the population. Thus, by "polypeptide" as used herein it is also meant to be inclusive of polypeptides encoded by an allelic variant of a nucleic acid molecule encoding a BSP. In a preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that encodes a polypeptide having the amino acid sequence selected from the group consisting of SEQ ID NO: 172 through 295. In a yet more preferred embodiment, the polypeptide is encoded by an allelic variant of a gene that has the nucleic acid sequence selected from the group consisting of SEQ ID NO: 1 through 171.

In another embodiment, the invention provides polypeptides which comprise derivatives of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred embodiment, the polypeptide is a BSP. In a preferred embodiment, the polypeptide has an amino acid sequence selected from the group consisting of SEQ ID NO: 172 through 295, or is a mutein, allelic variant, homologous protein or fragment thereof. In a preferred embodiment, the derivative has been acetylated, carboxylated, phosphorylated, glycosylated or ubiquitinated. In another preferred embodiment, the derivative has been labeled with, *e.g.*, radioactive isotopes such as ^{125}I, ^{32}P, ^{35}S, and ^{3}H. In another preferred embodiment, the derivative has been labeled with fluorophores, chemiluminescent agents, enzymes, and antiligands that can serve as specific binding pair members for a labeled ligand.

Polypeptide modifications are well-known to those of skill and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as, for instance Creighton, Protein Structure and Molecular Properties, 2nd ed., W. H. Freeman and Company (1993). Many detailed reviews are available on this subject, such as, for example, those provided by Wold, in Johnson (ed.), Posttranslational Covalent Modification of Proteins, pgs. 1-12, Academic Press (1983); Seifter *et al.*, *Meth. Enzymol.* 182: 626-646 (1990) and Rattan *et al.*, *Ann. N.Y. Acad. Sci.* 663: 48-62 (1992).

It will be appreciated, as is well-known and as noted above, that polypeptides are not always entirely linear. For instance, polypeptides may be branched as a result of ubiquitination, and they may be circular, with or without branching, generally as a result of posttranslation events, including natural processing event and events brought about by human manipulation which do not occur naturally. Circular, branched and branched circular polypeptides may be synthesized by non-translation natural process and by entirely synthetic methods, as well. Modifications can occur anywhere in a polypeptide, including the peptide backbone, the amino acid side-chains and the amino or carboxyl termini. In fact, blockage of the amino or carboxyl group in a polypeptide, or both, by a covalent modification, is common in naturally occurring and synthetic polypeptides and such modifications may be present in polypeptides of the present invention, as well. For instance, the amino terminal residue of polypeptides made in *E. coli*, prior to proteolytic processing, almost invariably will be N-formylmethionine.

Useful post-synthetic (and post-translational) modifications include conjugation to detectable labels, such as fluorophores. A wide variety of amine-reactive and thiol-reactive fluorophore derivatives have been synthesized that react under nondenaturing conditions with N-terminal amino groups and epsilon amino groups of lysine residues, on the one hand, and with free thiol groups of cysteine residues, on the other.

Kits are available commercially that permit conjugation of proteins to a variety of amine-reactive or thiol-reactive fluorophores: Molecular Probes, Inc. (Eugene, OR, USA), *e.g.*, offers kits for conjugating proteins to Alexa Fluor 350, Alexa Fluor 430, Fluorescein-EX, Alexa Fluor 488, Oregon Green 488, Alexa Fluor 532, Alexa Fluor 546, Alexa Fluor 546, Alexa Fluor 568, Alexa Fluor 594, and Texas Red-X.

A wide variety of other amine-reactive and thiol-reactive fluorophores are available commercially (Molecular Probes, Inc., Eugene, OR, USA), including Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA).

The polypeptides of the present invention can also be conjugated to fluorophores, other proteins, and other macromolecules, using bifunctional linking reagents. Common homobifunctional reagents include, *e.g.*, APG, AEDP, BASED, BMB, BMDB, BMH, BMOE, BM[PEO]3, BM[PEO]4, BS3, BSOCOES, DFDNB, DMA, DMP, DMS, DPDPB, DSG, DSP (Lomant's Reagent), DSS, DST, DTBP, DTME, DTSSP, EGS, HBVS, Sulfo-BSOCOES, Sulfo-DST, Sulfo-EGS (all available from Pierce, Rockford, IL, USA); common heterobifunctional cross-linkers include ABH, AMAS, ANB-NOS, APDP, ASBA, BMPA, BMPH, BMPS, EDC, EMCA, EMCH, EMCS, KMUA, KMUH, GMBS, LC-SMCC, LC-SPDP, MBS, M2C2H, MPBH, MSA, NHS-ASA, PDPH, PMPI, SADP, SAED, SAND, SANPAH, SASD, SATP, SBAP, SFAD, SIA, SIAB, SMCC, SMPB, SMPH, SMPT, SPDP, Sulfo-EMCS, Sulfo-GMBS, Sulfo-HSAB, Sulfo-KMUS, Sulfo-LC-SPDP, Sulfo-MBS, Sulfo-NHS-LC-ASA, Sulfo-SADP, Sulfo-SANPAH, Sulfo-SIAB, Sulfo-SMCC, Sulfo-SMPB, Sulfo-LC-SMPT, SVSB, TFCS (all available Pierce, Rockford, IL, USA).

The polypeptides, fragments, and fusion proteins of the present invention can be conjugated, using such cross-linking reagents, to fluorophores that are not amine- or thiol-reactive. Other labels that usefully can be conjugated to the polypeptides, fragments, and fusion proteins of the present invention include radioactive labels, echosonographic contrast reagents, and MRI contrast agents.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated using cross-linking agents to carrier proteins, such as KLH, bovine thyroglobulin, and even bovine serum albumin (BSA), to increase immunogenicity for raising anti-BSP antibodies.

The polypeptides, fragments, and fusion proteins of the present invention can also usefully be conjugated to polyethylene glycol (PEG); PEGylation increases the serum half-life of proteins administered intravenously for replacement therapy. Delgado *et al.*, *Crit. Rev. Ther. Drug Carrier Syst.* 9(3-4): 249-304 (1992); Scott *et al.*, *Curr. Pharm. Des.* 4(6): 423-38 (1998); DeSantis *et al.*, *Curr. Opin. Biotechnol.* 10(4): 324-30 (1999), incorporated herein by reference in their entireties. PEG monomers can be attached to the protein directly or through a linker, with PEGylation using PEG monomers activated with tresyl chloride (2,2,2-trifluoroethanesulphonyl chloride) permitting direct attachment under mild conditions.

In yet another embodiment, the invention provides analogs of a polypeptide encoded by a nucleic acid molecule according to the instant invention. In a preferred

embodiment, the polypeptide is a BSP. In a more preferred embodiment, the analog is derived from a polypeptide having part or all of the amino acid sequence of SEQ ID NO: 172 through 295. In a preferred embodiment, the analog is one that comprises one or more substitutions of non-natural amino acids or non-native inter-residue bonds compared to the naturally-occurring polypeptide. In general, the non-peptide analog is structurally similar to a BSP, but one or more peptide linkages is replaced by a linkage selected from the group consisting of --CH_2NH--, --CH_2S--, --CH_2-CH_2--, --CH=CH--(cis and trans), --$COCH_2$--, --$CH(OH)CH_2$-- and --CH_2SO--. In another embodiment, the non-peptide analog comprises substitution of one or more amino acids of a BSP with a D-amino acid of the same type or other non-natural amino acid in order to generate more stable peptides. D-amino acids can readily be incorporated during chemical peptide synthesis: peptides assembled from D-amino acids are more resistant to proteolytic attack; incorporation of D-amino acids can also be used to confer specific three-dimensional conformations on the peptide. Other amino acid analogues commonly added during chemical synthesis include ornithine, norleucine, phosphorylated amino acids (typically phosphoserine, phosphothreonine, phosphotyrosine), L-malonyltyrosine, a non-hydrolyzable analog of phosphotyrosine (*see, e.g.*, Kole *et al.*, *Biochem. Biophys. Res. Com.* 209: 817-821 (1995)), and various halogenated phenylalanine derivatives.

Non-natural amino acids can be incorporated during solid phase chemical synthesis or by recombinant techniques, although the former is typically more common. Solid phase chemical synthesis of peptides is well established in the art. Procedures are described, inter alia, in Chan *et al.* (eds.), Fmoc Solid Phase Peptide Synthesis: A Practical Approach (Practical Approach Series), Oxford Univ. Press (March 2000); Jones, Amino Acid and Peptide Synthesis (Oxford Chemistry Primers, No 7), Oxford Univ. Press (1992); and Bodanszky, Principles of Peptide Synthesis (Springer Laboratory), Springer Verlag (1993); the disclosures of which are incorporated herein by reference in their entireties.

Amino acid analogues having detectable labels are also usefully incorporated during synthesis to provide derivatives and analogs. Biotin, for example can be added using biotinoyl-(9-fluorenylmethoxycarbonyl)-L-lysine (FMOC biocytin) (Molecular Probes, Eugene, OR, USA). Biotin can also be added enzymatically by incorporation into a fusion protein of a *E. coli* BirA substrate peptide. The FMOC and *t*BOC derivatives of dabcyl-L-lysine (Molecular Probes, Inc., Eugene, OR, USA) can be used to incorporate the dabcyl chromophore at selected sites in the peptide sequence during

synthesis. The aminonaphthalene derivative EDANS, the most common fluorophore for pairing with the dabcyl quencher in fluorescence resonance energy transfer (FRET) systems, can be introduced during automated synthesis of peptides by using EDANS-FMOC-L-glutamic acid or the corresponding *t*BOC derivative (both from Molecular Probes, Inc., Eugene, OR, USA). Tetramethylrhodamine fluorophores can be incorporated during automated FMOC synthesis of peptides using (FMOC)-TMR-L-lysine (Molecular Probes, Inc. Eugene, OR, USA).

Other useful amino acid analogues that can be incorporated during chemical synthesis include aspartic acid, glutamic acid, lysine, and tyrosine analogues having allyl side-chain protection (Applied Biosystems, Inc., Foster City, CA, USA); the allyl side chain permits synthesis of cyclic, branched-chain, sulfonated, glycosylated, and phosphorylated peptides.

A large number of other FMOC-protected non-natural amino acid analogues capable of incorporation during chemical synthesis are available commercially, including, *e.g.*, Fmoc-2-aminobicyclo[2.2.1]heptane-2-carboxylic acid, Fmoc-3-endo-aminobicyclo[2.2.1]heptane-2-endo-carboxylic acid, Fmoc-3-exo-aminobicyclo[2.2.1]heptane-2-exo-carboxylic acid, Fmoc-3-endo-amino-bicyclo[2.2.1]hept-5-ene-2-endo-carboxylic acid, Fmoc-3-exo-amino-bicyclo[2.2.1]hept-5-ene-2-exo-carboxylic acid, Fmoc-cis-2-amino-1-cyclohexanecarboxylic acid, Fmoc-trans-2-amino-1-cyclohexanecarboxylic acid, Fmoc-1-amino-1-cyclopentanecarboxylic acid, Fmoc-cis-2-amino-1-cyclopentanecarboxylic acid, Fmoc-1-amino-1-cyclopropanecarboxylic acid, Fmoc-D-2-amino-4-(ethylthio)butyric acid, Fmoc-L-2-amino-4-(ethylthio)butyric acid, Fmoc-L-buthionine, Fmoc-S-methyl-L-Cysteine, Fmoc-2-aminobenzoic acid (anthranillic acid), Fmoc-3-aminobenzoic acid, Fmoc-4-aminobenzoic acid, Fmoc-2-aminobenzophenone-2'-carboxylic acid, Fmoc-N-(4-aminobenzoyl)-β-alanine, Fmoc-2-amino-4,5-dimethoxybenzoic acid, Fmoc-4-aminohippuric acid, Fmoc-2-amino-3-hydroxybenzoic acid, Fmoc-2-amino-5-hydroxybenzoic acid, Fmoc-3-amino-4-hydroxybenzoic acid, Fmoc-4-amino-3-hydroxybenzoic acid, Fmoc-4-amino-2-hydroxybenzoic acid, Fmoc-5-amino-2-hydroxybenzoic acid, Fmoc-2-amino-3-methoxybenzoic acid, Fmoc-4-amino-3-methoxybenzoic acid, Fmoc-2-amino-3-methylbenzoic acid, Fmoc-2-amino-5-methylbenzoic acid, Fmoc-2-amino-6-methylbenzoic acid, Fmoc-3-amino-2-methylbenzoic acid, Fmoc-3-amino-4-methylbenzoic acid, Fmoc-4-amino-3-methylbenzoic acid, Fmoc-3-amino-2-naphtoic acid, Fmoc-D,L-3-amino-3-

phenylpropionic acid, Fmoc-L-Methyldopa, Fmoc-2-amino-4,6-dimethyl-3-pyridinecarboxylic acid, Fmoc-D,L-amino-2-thiophenacetic acid, Fmoc-4-(carboxymethyl)piperazine, Fmoc-4-carboxypiperazine, Fmoc-4-(carboxymethyl)homopiperazine, Fmoc-4-phenyl-4-piperidinecarboxylic acid, Fmoc-L-1,2,3,4-tetrahydronorharman-3-carboxylic acid, Fmoc-L-thiazolidine-4-carboxylic acid, all available from The Peptide Laboratory (Richmond, CA, USA).

Non-natural residues can also be added biosynthetically by engineering a suppressor tRNA, typically one that recognizes the UAG stop codon, by chemical aminoacylation with the desired unnatural amino acid. Conventional site-directed mutagenesis is used to introduce the chosen stop codon UAG at the site of interest in the protein gene. When the acylated suppressor tRNA and the mutant gene are combined in an *in vitro* transcription/translation system, the unnatural amino acid is incorporated in response to the UAG codon to give a protein containing that amino acid at the specified position. Liu *et al.*, *Proc. Natl Acad. Sci. USA* 96(9): 4780-5 (1999); Wang *et al.*, *Science* 292(5516): 498-500 (2001).

Fusion Proteins

The present invention further provides fusions of each of the polypeptides and fragments of the present invention to heterologous polypeptides. In a preferred embodiment, the polypeptide is a BSP. In a more preferred embodiment, the polypeptide that is fused to the heterologous polypeptide comprises part or all of the amino acid sequence of SEQ ID NO: 172 through 295, or is a mutein, homologous polypeptide, analog or derivative thereof. In an even more preferred embodiment, the nucleic acid molecule encoding the fusion protein comprises all or part of the nucleic acid sequence of SEQ ID NO: 1 through 171, or comprises all or part of a nucleic acid sequence that selectively hybridizes or is homologous to a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 171.

The fusion proteins of the present invention will include at least one fragment of the protein of the present invention, which fragment is at least 6, typically at least 8, often at least 15, and usefully at least 16, 17, 18, 19, or 20 amino acids long. The fragment of the protein of the present to be included in the fusion can usefully be at least 25 amino acids long, at least 50 amino acids long, and can be at least 75, 100, or even 150 amino acids long. Fusions that include the entirety of the proteins of the present invention have particular utility.

The heterologous polypeptide included within the fusion protein of the present invention is at least 6 amino acids in length, often at least 8 amino acids in length, and usefully at least 15, 20, and 25 amino acids in length. Fusions that include larger polypeptides, such as the IgG Fc region, and even entire proteins (such as GFP chromophore-containing proteins) are particular useful.

As described above in the description of vectors and expression vectors of the present invention, which discussion is incorporated here by reference in its entirety, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those designed to facilitate purification and/or visualization of recombinantly-expressed proteins. *See, e.g.*, Ausubel, Chapter 16, (1992), *supra*. Although purification tags can also be incorporated into fusions that are chemically synthesized, chemical synthesis typically provides sufficient purity that further purification by HPLC suffices; however, visualization tags as above described retain their utility even when the protein is produced by chemical synthesis, and when so included render the fusion proteins of the present invention useful as directly detectable markers of the presence of a polypeptide of the invention.

As also discussed above, heterologous polypeptides to be included in the fusion proteins of the present invention can usefully include those that facilitate secretion of recombinantly expressed proteins — into the periplasmic space or extracellular milieu for prokaryotic hosts, into the culture medium for eukaryotic cells — through incorporation of secretion signals and/or leader sequences. For example, a His^6 tagged protein can be purified on a Ni affinity column and a GST fusion protein can be purified on a glutathione affinity column. Similarly, a fusion protein comprising the Fc domain of IgG can be purified on a Protein A or Protein G column and a fusion protein comprising an epitope tag such as myc can be purified using an immunoaffinity column containing an anti-c-myc antibody. It is preferable that the epitope tag be separated from the protein encoded by the essential gene by an enzymatic cleavage site that can be cleaved after purification. See also the discussion of nucleic acid molecules encoding fusion proteins that may be expressed on the surface of a cell.

Other useful protein fusions of the present invention include those that permit use of the protein of the present invention as bait in a yeast two-hybrid system. *See* Bartel *et al.* (eds.), The Yeast Two-Hybrid System, Oxford University Press (1997); Zhu *et al.*, Yeast Hybrid Technologies, Eaton Publishing (2000); Fields *et al.*, *Trends Genet.* 10(8): 286-92 (1994); Mendelsohn *et al.*, *Curr. Opin. Biotechnol.* 5(5): 482-6 (1994); Luban *et*

al., *Curr. Opin. Biotechnol.* 6(1): 59-64 (1995); Allen *et al.*, *Trends Biochem. Sci.* 20(12): 511-6 (1995); Drees, *Curr. Opin. Chem. Biol.* 3(1): 64-70 (1999); Topcu *et al.*, *Pharm. Res.* 17(9): 1049-55 (2000); Fashena *et al.*, *Gene* 250(1-2): 1-14 (2000); ; Colas *et al.*, (1996) Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380, 548-550; Norman, T. *et al.*, (1999) Genetic selection of peptide inhibitors of biological pathways. *Science* 285, 591-595, Fabbrizio *et al.*, (1999) Inhibition of mammalian cell proliferation by genetically selected peptide aptamers that functionally antagonize E2F activity. *Oncogene* 18, 4357-4363; Xu *et al.*, (1997) Cells that register logical relationships among proteins. *Proc Natl Acad Sci U S A.* 94, 12473-12478; Yang, *et al.*, (1995) Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nuc. Acids Res.* 23, 1152-1156; Kolonin *et al.*, (1998) Targeting cyclin-dependent kinases in Drosophila with peptide aptamers. *Proc Natl Acad Sci U S A* 95, 14266-14271; Cohen *et al.*, (1998) An artificial cell-cycle inhibitor isolated from a combinatorial library. *Proc Natl Acad Sci U S A* 95, 14272-14277; Uetz, P.; Giot, L.; al, e.; Fields, S.; Rothberg, J. M. (2000) A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae. *Nature* 403, 623-627; Ito, *et al.*, (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci U S A* 98, 4569-4574, the disclosures of which are incorporated herein by reference in their entireties. Typically, such fusion is to either *E. coli* LexA or yeast GAL4 DNA binding domains. Related bait plasmids are available that express the bait fused to a nuclear localization signal.

Other useful fusion proteins include those that permit display of the encoded protein on the surface of a phage or cell, fusions to intrinsically fluorescent proteins, such as green fluorescent protein (GFP), and fusions to the IgG Fc region, as described above, which discussion is incorporated here by reference in its entirety.

The polypeptides and fragments of the present invention can also usefully be fused to protein toxins, such as *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, ricin, in order to effect ablation of cells that bind or take up the proteins of the present invention.

Fusion partners include, *inter alia*, *myc*, hemagglutinin (HA), GST, immunoglobulins, β-galactosidase, biotin trpE, protein A, β-lactamase, α-amylase, maltose binding protein, alcohol dehydrogenase, polyhistidine (for example, six histidine at the amino and/or carboxyl terminus of the polypeptide), lacZ, green fluorescent protein (GFP), yeast $\underline{\alpha}$ mating factor, GAL4 transcription activation or DNA binding domain,

luciferase, and serum proteins such as ovalbumin, albumin and the constant domain of IgG. *See, e.g.*, Ausubel (1992), *supra* and Ausubel (1999), *supra*. Fusion proteins may also contain sites for specific enzymatic cleavage, such as a site that is recognized by enzymes such as Factor XIII, trypsin, pepsin, or any other enzyme known in the art. Fusion proteins will typically be made by either recombinant nucleic acid methods, as described above, chemically synthesized using techniques well-known in the art (*e.g.*, a Merrifield synthesis), or produced by chemical cross-linking.

Another advantage of fusion proteins is that the epitope tag can be used to bind the fusion protein to a plate or column through an affinity linkage for screening binding proteins or other molecules that bind to the BSP.

As further described below, the isolated polypeptides, muteins, fusion proteins, homologous proteins or allelic variants of the present invention can readily be used as specific immunogens to raise antibodies that specifically recognize BSPs, their allelic variants and homologues. The antibodies, in turn, can be used, *inter alia*, specifically to assay for the polypeptides of the present invention, particularly BSPs, *e.g.* by ELISA for detection of protein fluid samples, such as serum, by immunohistochemistry or laser scanning cytometry, for detection of protein in tissue samples, or by flow cytometry, for detection of intracellular protein in cell suspensions, for specific antibody-mediated isolation and/or purification of BSPs, as for example by immunoprecipitation, and for use as specific agonists or antagonists of BSPs.

One may determine whether polypeptides including muteins, fusion proteins, homologous proteins or allelic variants are functional by methods known in the art. For instance, residues that are tolerant of change while retaining function can be identified by altering the protein at known residues using methods known in the art, such as alanine scanning mutagenesis, Cunningham *et al.*, *Science* 244(4908): 1081-5 (1989); transposon linker scanning mutagenesis, Chen *et al.*, *Gene* 263(1-2): 39-48 (2001); combinations of homolog- and alanine-scanning mutagenesis, Jin *et al.*, *J. Mol. Biol.* 226(3): 851-65 (1992); combinatorial alanine scanning, Weiss *et al.*, *Proc. Natl. Acad. Sci USA* 97(16): 8950-4 (2000), followed by functional assay. Transposon linker scanning kits are available commercially (New England Biolabs, Beverly, MA, USA, catalog. no. E7-102S; EZ::TN™ In-Frame Linker Insertion Kit, catalogue no. EZI04KN, Epicentre Technologies Corporation, Madison, WI, USA).

Purification of the polypeptides including fragments, homologous polypeptides, muteins, analogs, derivatives and fusion proteins is well-known and within the skill of

one having ordinary skill in the art. *See, e.g.*, Scopes, Protein Purification, 2d ed. (1987). Purification of recombinantly expressed polypeptides is described above. Purification of chemically-synthesized peptides can readily be effected, *e.g.*, by HPLC.

Accordingly, it is an aspect of the present invention to provide the isolated proteins of the present invention in pure or substantially pure form in the presence of absence of a stabilizing agent. Stabilizing agents include both proteinaceous or non-proteinaceous material and are well-known in the art. Stabilizing agents, such as albumin and polyethylene glycol (PEG) are known and are commercially available.

Although high levels of purity are preferred when the isolated proteins of the present invention are used as therapeutic agents, such as in vaccines and as replacement therapy, the isolated proteins of the present invention are also useful at lower purity. For example, partially purified proteins of the present invention can be used as immunogens to raise antibodies in laboratory animals.

In preferred embodiments, the purified and substantially purified proteins of the present invention are in compositions that lack detectable ampholytes, acrylamide monomers, bis-acrylamide monomers, and polyacrylamide.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be attached to a substrate. The substrate can be porous or solid, planar or non-planar; the bond can be covalent or noncovalent.

For example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a porous substrate, commonly a membrane, typically comprising nitrocellulose, polyvinylidene fluoride (PVDF), or cationically derivatized, hydrophilic PVDF; so bound, the proteins, fragments, and fusions of the present invention can be used to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention.

As another example, the polypeptides, fragments, analogs, derivatives and fusions of the present invention can usefully be bound to a substantially nonporous substrate, such as plastic, to detect and quantify antibodies, *e.g.* in serum, that bind specifically to the immobilized protein of the present invention. Such plastics include polymethylacrylic, polyethylene, polypropylene, polyacrylate, polymethylmethacrylate, polyvinylchloride, polytetrafluoroethylene, polystyrene, polycarbonate, polyacetal, polysulfone, celluloseacetate, cellulosenitrate, nitrocellulose, or mixtures thereof; when the assay is performed in a standard microtiter dish, the plastic is typically polystyrene.

The polypeptides, fragments, analogs, derivatives and fusions of the present invention can also be attached to a substrate suitable for use as a surface enhanced laser desorption ionization source; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biologic interaction there between. The proteins, fragments, and fusions of the present invention can also be attached to a substrate suitable for use in surface plasmon resonance detection; so attached, the protein, fragment, or fusion of the present invention is useful for binding and then detecting secondary proteins that bind with sufficient affinity or avidity to the surface-bound protein to indicate biological interaction there between.

Antibodies

In another aspect, the invention provides antibodies, including fragments and derivatives thereof, that bind specifically to polypeptides encoded by the nucleic acid molecules of the invention, as well as antibodies that bind to fragments, muteins, derivatives and analogs of the polypeptides. In a preferred embodiment, the antibodies are specific for a polypeptide that is a BSP, or a fragment, mutein, derivative, analog or fusion protein thereof. In a more preferred embodiment, the antibodies are specific for a polypeptide that comprises SEQ ID NO: 172 through 295, or a fragment, mutein, derivative, analog or fusion protein thereof.

The antibodies of the present invention can be specific for linear epitopes, discontinuous epitopes, or conformational epitopes of such proteins or protein fragments, either as present on the protein in its native conformation or, in some cases, as present on the proteins as denatured, as, *e.g.*, by solubilization in SDS. New epitopes may be also due to a difference in post translational modifications (PTMs) in disease versus normal tissue. For example, a particular site on a BSP may be glycosylated in cancerous cells, but not glycosylated in normal cells or visa versa. In addition, alternative splice forms of a BSP may be indicative of cancer. Differential degradation of the C or N-terminus of a BSP may also be a marker or target for anticancer therapy. For example, a BSP may be N-terminal degraded in cancer cells exposing new epitopes to which antibodies may selectively bind for diagnostic or therapeutic uses.

As is well-known in the art, the degree to which an antibody can discriminate as among molecular species in a mixture will depend, in part, upon the conformational relatedness of the species in the mixture; typically, the antibodies of the present invention

will discriminate over adventitious binding to non-BSP polypeptides by at least 2-fold, more typically by at least 5-fold, typically by more than 10-fold, 25-fold, 50-fold, 75-fold, and often by more than 100-fold, and on occasion by more than 500-fold or 1000-fold. When used to detect the proteins or protein fragments of the present invention, the antibody of the present invention is sufficiently specific when it can be used to determine the presence of the protein of the present invention in samples derived from human breast.

Typically, the affinity or avidity of an antibody (or antibody multimer, as in the case of an IgM pentamer) of the present invention for a protein or protein fragment of the present invention will be at least about 1×10^{-6} molar (M), typically at least about 5×10^{-7} M, 1×10^{-7} M, with affinities and avidities of at least 1×10^{-8} M, 5×10^{-9} M, 1×10^{-10} M and up to 1×10^{-13} M proving especially useful.

The antibodies of the present invention can be naturally-occurring forms, such as IgG, IgM, IgD, IgE, IgY, and IgA, from any avian, reptilian, or mammalian species.

Human antibodies can, but will infrequently, be drawn directly from human donors or human cells. In this case, antibodies to the proteins of the present invention will typically have resulted from fortuitous immunization, such as autoimmune immunization, with the protein or protein fragments of the present invention. Such antibodies will typically, but will not invariably, be polyclonal. In addition, individual polyclonal antibodies may be isolated and cloned to generate monoclonals.

Human antibodies are more frequently obtained using transgenic animals that express human immunoglobulin genes, which transgenic animals can be affirmatively immunized with the protein immunogen of the present invention. Human Ig-transgenic mice capable of producing human antibodies and methods of producing human antibodies therefrom upon specific immunization are described, *inter alia*, in U.S. Patents 6,162,963; 6,150,584; 6,114,598; 6,075,181; 5,939,598; 5,877,397; 5,874,299; 5,814,318; 5,789,650; 5,770,429; 5,661,016; 5,633,425; 5,625,126; 5,569,825; 5,545,807; 5,545,806, and 5,591,669, the disclosures of which are incorporated herein by reference in their entireties. Such antibodies are typically monoclonal, and are typically produced using techniques developed for production of murine antibodies.

Human antibodies are particularly useful, and often preferred, when the antibodies of the present invention are to be administered to human beings as *in vivo* diagnostic or therapeutic agents, since recipient immune response to the administered

antibody will often be substantially less than that occasioned by administration of an antibody derived from another species, such as mouse.

IgG, IgM, IgD, IgE, IgY, and IgA antibodies of the present invention can also be obtained from other species, including mammals such as rodents (typically mouse, but also rat, guinea pig, and hamster) lagomorphs, typically rabbits, and also larger mammals, such as sheep, goats, cows, and horses, and other egg laying birds or reptiles such as chickens or alligators. For example, avian antibodies may be generated using techniques described in WO 00/29444, published 25 May 2000, the contents of which are hereby incorporated in their entirety. In such cases, as with the transgenic human-antibody-producing non-human mammals, fortuitous immunization is not required, and the non-human mammal is typically affirmatively immunized, according to standard immunization protocols, with the protein or protein fragment of the present invention.

As discussed above, virtually all fragments of 8 or more contiguous amino acids of the proteins of the present invention can be used effectively as immunogens when conjugated to a carrier, typically a protein such as bovine thyroglobulin, keyhole limpet hemocyanin, or bovine serum albumin, conveniently using a bifunctional linker such as those described elsewhere above, which discussion is incorporated by reference here.

Immunogenicity can also be conferred by fusion of the polypeptide and fragments of the present invention to other moieties. For example, peptides of the present invention can be produced by solid phase synthesis on a branched polylysine core matrix; these multiple antigenic peptides (MAPs) provide high purity, increased avidity, accurate chemical definition and improved safety in vaccine development. Tam *et al.*, *Proc. Natl. Acad. Sci. USA* 85: 5409-5413 (1988); Posnett *et al.*, *J. Biol. Chem.* 263: 1719-1725 (1988).

Protocols for immunizing non-human mammals or avian species are well-established in the art. *See* Harlow *et al.* (eds.), Using Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory (1998); Coligan *et al.* (eds.), Current Protocols in Immunology, John Wiley & Sons, Inc. (2001); Zola, Monoclonal Antibodies: Preparation and Use of Monoclonal Antibodies and Engineered Antibody Derivatives (Basics: From Background to Bench), Springer Verlag (2000); Gross M, Speck *J.Dtsch. Tierarztl. Wochenschr.* 103: 417-422 (1996), the disclosures of which are incorporated herein by reference. Immunization protocols often include multiple immunizations, either with or without adjuvants such as Freund's complete adjuvant and Freund's incomplete adjuvant, and may include naked DNA immunization (Moss, *Semin. Immunol.* 2: 317-327 (1990).

Antibodies from non-human mammals and avian species can be polyclonal or monoclonal, with polyclonal antibodies having certain advantages in immunohistochemical detection of the proteins of the present invention and monoclonal antibodies having advantages in identifying and distinguishing particular epitopes of the proteins of the present invention. Antibodies from avian species may have particular advantage in detection of the proteins of the present invention, in human serum or tissues (Vikinge et al., *Biosens. Bioelectron.* 13: 1257-1262 (1998).

Following immunization, the antibodies of the present invention can be produced using any art-accepted technique. Such techniques are well-known in the art, Coligan, *supra*; Zola, *supra*; Howard *et al.* (eds.), Basic Methods in Antibody Production and Characterization, CRC Press (2000); Harlow, *supra*; Davis (ed.), Monoclonal Antibody Protocols, Vol. 45, Humana Press (1995); Delves (ed.), Antibody Production: Essential Techniques, John Wiley & Son Ltd (1997); Kenney, Antibody Solution: An Antibody Methods Manual, Chapman & Hall (1997), incorporated herein by reference in their entireties, and thus need not be detailed here.

Briefly, however, such techniques include, *inter alia*, production of monoclonal antibodies by hybridomas and expression of antibodies or fragments or derivatives thereof from host cells engineered to express immunoglobulin genes or fragments thereof. These two methods of production are not mutually exclusive: genes encoding antibodies specific for the proteins or protein fragments of the present invention can be cloned from hybridomas and thereafter expressed in other host cells. Nor need the two necessarily be performed together: *e.g.*, genes encoding antibodies specific for the proteins and protein fragments of the present invention can be cloned directly from B cells known to be specific for the desired protein, as further described in U.S Patent 5,627,052, the disclosure of which is incorporated herein by reference in its entirety, or from antibody-displaying phage.

Recombinant expression in host cells is particularly useful when fragments or derivatives of the antibodies of the present invention are desired.

Host cells for recombinant production of either whole antibodies, antibody fragments, or antibody derivatives can be prokaryotic or eukaryotic.

Prokaryotic hosts are particularly useful for producing phage displayed antibodies of the present invention.

The technology of phage-displayed antibodies, in which antibody variable region fragments are fused, for example, to the gene III protein (pIII) or gene VIII protein

(pVIII) for display on the surface of filamentous phage, such as M13, is by now well-established. *See, e.g.,* Sidhu, *Curr. Opin. Biotechnol.* 11(6): 610-6 (2000); Griffiths *et al., Curr. Opin. Biotechnol.* 9(1): 102-8 (1998); Hoogenboom *et al., Immunotechnology,* 4(1): 1-20 (1998); Rader *et al., Current Opinion in Biotechnology* 8: 503-508 (1997); Aujame *et al., Human Antibodies* 8: 155-168 (1997); Hoogenboom, *Trends in Biotechnol.* 15: 62-70 (1997); de Kruif *et al.*, 17: 453-455 (1996); Barbas *et al., Trends in Biotechnol.* 14: 230-234 (1996); Winter *et al., Ann. Rev. Immunol.* 433-455 (1994). Techniques and protocols required to generate, propagate, screen (pan), and use the antibody fragments from such libraries have recently been compiled. *See, e.g.,* Barbas (2001), *supra*; Kay, *supra*; Abelson, *supra*, the disclosures of which are incorporated herein by reference in their entireties.

Typically, phage-displayed antibody fragments are scFv fragments or Fab fragments; when desired, full length antibodies can be produced by cloning the variable regions from the displaying phage into a complete antibody and expressing the full length antibody in a further prokaryotic or a eukaryotic host cell.

Eukaryotic cells are also useful for expression of the antibodies, antibody fragments, and antibody derivatives of the present invention.

For example, antibody fragments of the present invention can be produced in *Pichia pastoris* and in *Saccharomyces cerevisiae*. *See, e.g.,* Takahashi *et al., Biosci. Biotechnol. Biochem.* 64(10): 2138-44 (2000); Freyre *et al.*, J. Biotechnol. 76(2-3):1 57-63 (2000); Fischer *et al., Biotechnol. Appl. Biochem.* 30 (Pt 2): 117-20 (1999); Pennell *et al., Res. Immunol.* 149(6): 599-603 (1998); Eldin *et al., J. Immunol. Methods.* 201(1): 67-75 (1997);, Frenken *et al., Res. Immunol.* 149(6): 589-99 (1998); Shusta *et al., Nature Biotechnol.* 16(8): 773-7 (1998), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in insect cells. *See, e.g.,* Li *et al., Protein Expr. Purif.* 21(1): 121-8 (2001); Ailor *et al., Biotechnol. Bioeng.* 58(2-3): 196-203 (1998); Hsu *et al., Biotechnol. Prog.* 13(1): 96-104 (1997); Edelman *et al., Immunology* 91(1): 13-9 (1997); and Nesbit *et al., J. Immunol. Methods* 151(1-2): 201-8 (1992), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies and fragments and derivatives thereof of the present invention can also be produced in plant cells, particularly maize or tobacco, Giddings *et al., Nature Biotechnol.* 18(11): 1151-5 (2000); Gavilondo *et al., Biotechniques* 29(1): 128-38 (2000);

Fischer *et al.*, *J. Biol. Regul. Homeost. Agents* 14(2): 83-92 (2000); Fischer *et al.*, *Biotechnol. Appl. Biochem.* 30 (Pt 2): 113-6 (1999); Fischer *et al.*, *Biol. Chem.* 380(7-8): 825-39 (1999); Russell, *Curr. Top. Microbiol. Immunol.* 240: 119-38 (1999); and Ma *et al.*, *Plant Physiol.* 109(2): 341-6 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Antibodies, including antibody fragments and derivatives, of the present invention can also be produced in transgenic, non-human, mammalian milk. *See, e.g.* Pollock et al., *J. Immunol Methods.* 231: 147-57 (1999); Young et al., *Res. Immunol.* 149: 609-10 (1998); Limonta et al., *Immunotechnology* 1: 107-13 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Mammalian cells useful for recombinant expression of antibodies, antibody fragments, and antibody derivatives of the present invention include CHO cells, COS cells, 293 cells, and myeloma cells.

Verma *et al.*, *J. Immunol. Methods* 216(1-2):165-81 (1998), herein incorporated by reference, review and compare bacterial, yeast, insect and mammalian expression systems for expression of antibodies.

Antibodies of the present invention can also be prepared by cell free translation, as further described in Merk *et al.*, *J. Biochem.* (Tokyo) 125(2): 328-33 (1999) and Ryabova *et al.*, *Nature Biotechnol.* 15(1): 79-84 (1997), and in the milk of transgenic animals, as further described in Pollock *et al.*, *J. Immunol. Methods* 231(1-2): 147-57 (1999), the disclosures of which are incorporated herein by reference in their entireties.

The invention further provides antibody fragments that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful fragments are Fab, Fab', Fv, $F(ab)'_2$, and single chain Fv (scFv) fragments. Other useful fragments are described in Hudson, *Curr. Opin. Biotechnol.* 9(4): 395-402 (1998).

It is also an aspect of the present invention to provide antibody derivatives that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively

inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

Among such useful derivatives are chimeric, primatized, and humanized antibodies; such derivatives are less immunogenic in human beings, and thus more suitable for *in vivo* administration, than are unmodified antibodies from non-human mammalian species. Another useful derivative is PEGylation to increase the serum half life of the antibodies.

Chimeric antibodies typically include heavy and/or light chain variable regions (including both CDR and framework residues) of immunoglobulins of one species, typically mouse, fused to constant regions of another species, typically human. *See, e.g.*, United States Patent No. 5,807,715; Morrison *et al.*, *Proc. Natl. Acad. Sci USA*.81(21): 6851-5 (1984); Sharon *et al.*, *Nature* 309(5966): 364-7 (1984); Takeda *et al.*, *Nature* 314(6010): 452-4 (1985), the disclosures of which are incorporated herein by reference in their entireties. Primatized and humanized antibodies typically include heavy and/or light chain CDRs from a murine antibody grafted into a non-human primate or human antibody V region framework, usually further comprising a human constant region, Riechmann *et al.*, *Nature* 332(6162): 323-7 (1988); Co *et al.*, *Nature* 351(6326): 501-2 (1991); United States Patent Nos. 6,054,297; 5,821,337; 5,770,196; 5,766,886; 5,821,123; 5,869,619; 6,180,377; 6,013,256; 5,693,761; and 6,180,370, the disclosures of which are incorporated herein by reference in their entireties.

Other useful antibody derivatives of the invention include heteromeric antibody complexes and antibody fusions, such as diabodies (bispecific antibodies), single-chain diabodies, and intrabodies.

It is contemplated that the nucleic acids encoding the antibodies of the present invention can be operably joined to other nucleic acids forming a recombinant vector for cloning or for expression of the antibodies of the invention. The present invention includes any recombinant vector containing the coding sequences, or part thereof, whether for eukaryotic transduction, transfection or gene therapy. Such vectors may be prepared using conventional molecular biology techniques, known to those with skill in the art, and would comprise DNA encoding sequences for the immunoglobulin V-regions including framework and CDRs or parts thereof, and a suitable promoter either with or without a signal sequence for intracellular transport. Such vectors may be transduced or transfected into eukaryotic cells or used for gene therapy (Marasco et al., <u>Proc. Natl.</u>

Acad. Sci. (USA) 90: 7889-7893 (1993); Duan et al., Proc. Natl. Acad. Sci. (USA) 91: 5075-5079 (1994), by conventional techniques, known to those with skill in the art.

The antibodies of the present invention, including fragments and derivatives thereof, can usefully be labeled. It is, therefore, another aspect of the present invention to provide labeled antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

The choice of label depends, in part, upon the desired use.

For example, when the antibodies of the present invention are used for immunohistochemical staining of tissue samples, the label is preferably an enzyme that catalyzes production and local deposition of a detectable product.

Enzymes typically conjugated to antibodies to permit their immunohistochemical visualization are well-known, and include alkaline phosphatase, β-galactosidase, glucose oxidase, horseradish peroxidase (HRP), and urease. Typical substrates for production and deposition of visually detectable products include o-nitrophenyl-beta-D-galactopyranoside (ONPG); o-phenylenediamine dihydrochloride (OPD); p-nitrophenyl phosphate (PNPP); p-nitrophenyl-beta-D-galactopryanoside (PNPG); 3',3'-diaminobenzidine (DAB); 3-amino-9-ethylcarbazole (AEC); 4-chloro-1-naphthol (CN); 5-bromo-4-chloro-3-indolyl-phosphate (BCIP); ABTS®; BluoGal; iodonitrotetrazolium (INT); nitroblue tetrazolium chloride (NBT); phenazine methosulfate (PMS); phenolphthalein monophosphate (PMP); tetramethyl benzidine (TMB); tetranitroblue tetrazolium (TNBT); X-Gal; X-Gluc; and X-Glucoside.

Other substrates can be used to produce products for local deposition that are luminescent. For example, in the presence of hydrogen peroxide (H_2O_2), horseradish peroxidase (HRP) can catalyze the oxidation of cyclic diacylhydrazides, such as luminol. Immediately following the oxidation, the luminol is in an excited state (intermediate reaction product), which decays to the ground state by emitting light. Strong enhancement of the light emission is produced by enhancers, such as phenolic compounds. Advantages include high sensitivity, high resolution, and rapid detection without radioactivity and requiring only small amounts of antibody. *See, e.g.*, Thorpe *et al.*, *Methods Enzymol.* 133: 331-53 (1986); Kricka *et al.*, *J. Immunoassay* 17(1): 67-83

(1996); and Lundqvist *et al.*, *J. Biolumin. Chemilumin.* 10(6): 353-9 (1995), the disclosures of which are incorporated herein by reference in their entireties. Kits for such enhanced chemiluminescent detection (ECL) are available commercially.

The antibodies can also be labeled using colloidal gold.

As another example, when the antibodies of the present invention are used, *e.g.*, for flow cytometric detection, for scanning laser cytometric detection, or for fluorescent immunoassay, they can usefully be labeled with fluorophores.

There are a wide variety of fluorophore labels that can usefully be attached to the antibodies of the present invention.

For flow cytometric applications, both for extracellular detection and for intracellular detection, common useful fluorophores can be fluorescein isothiocyanate (FITC), allophycocyanin (APC), R-phycoerythrin (PE), peridinin chlorophyll protein (PerCP), Texas Red, Cy3, Cy5, fluorescence resonance energy tandem fluorophores such as PerCP-Cy5.5, PE-Cy5, PE-Cy5.5, PE-Cy7, PE-Texas Red, and APC-Cy7.

Other fluorophores include, *inter alia*, Alexa Fluor® 350, Alexa Fluor® 488, Alexa Fluor® 532, Alexa Fluor® 546, Alexa Fluor® 568, Alexa Fluor® 594, Alexa Fluor® 647 (monoclonal antibody labeling kits available from Molecular Probes, Inc., Eugene, OR, USA), BODIPY dyes, such as BODIPY 493/503, BODIPY FL, BODIPY R6G, BODIPY 530/550, BODIPY TMR, BODIPY 558/568, BODIPY 558/568, BODIPY 564/570, BODIPY 576/589, BODIPY 581/591, BODIPY TR, BODIPY 630/650, BODIPY 650/665, Cascade Blue, Cascade Yellow, Dansyl, lissamine rhodamine B, Marina Blue, Oregon Green 488, Oregon Green 514, Pacific Blue, rhodamine 6G, rhodamine green, rhodamine red, tetramethylrhodamine, Texas Red (available from Molecular Probes, Inc., Eugene, OR, USA), and Cy2, Cy3, Cy3.5, Cy5, Cy5.5, Cy7, all of which are also useful for fluorescently labeling the antibodies of the present invention.

For secondary detection using labeled avidin, streptavidin, captavidin or neutravidin, the antibodies of the present invention can usefully be labeled with biotin.

When the antibodies of the present invention are used, *e.g.*, for Western blotting applications, they can usefully be labeled with radioisotopes, such as ^{33}P, ^{32}P, ^{35}S, ^{3}H, and ^{125}I.

As another example, when the antibodies of the present invention are used for radioimmunotherapy, the label can usefully be ^{228}Th, ^{227}Ac, ^{225}Ac, ^{223}Ra, ^{213}Bi, ^{212}Pb,

^{212}Bi, ^{211}At, ^{203}Pb, ^{194}Os, ^{188}Re, ^{186}Re, ^{153}Sm, ^{149}Tb, ^{131}I, ^{125}I, ^{111}In, ^{105}Rh, ^{99m}Tc, ^{97}Ru, ^{90}Y, ^{90}Sr, ^{88}Y, ^{72}Se, ^{67}Cu, or ^{47}Sc.

As another example, when the antibodies of the present invention are to be used for *in vivo* diagnostic use, they can be rendered detectable by conjugation to MRI contrast agents, such as gadolinium diethylenetriaminepentaacetic acid (DTPA), Lauffer *et al.*, *Radiology* 207(2): 529-38 (1998), or by radioisotopic labeling.

As would be understood, use of the labels described above is not restricted to the application for which they are mentioned.

The antibodies of the present invention, including fragments and derivatives thereof, can also be conjugated to toxins, in order to target the toxin's ablative action to cells that display and/or express the proteins of the present invention. Commonly, the antibody in such immunotoxins is conjugated to *Pseudomonas* exotoxin A, *diphtheria* toxin, *shiga* toxin A, *anthrax* toxin lethal factor, or ricin. *See* Hall (ed.), Immunotoxin Methods and Protocols (Methods in Molecular Biology, vol. 166), Humana Press (2000); and Frankel *et al.* (eds.), Clinical Applications of Immunotoxins, Springer-Verlag (1998), the disclosures of which are incorporated herein by reference in their entireties.

The antibodies of the present invention can usefully be attached to a substrate, and it is, therefore, another aspect of the invention to provide antibodies that bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, attached to a substrate.

Substrates can be porous or nonporous, planar or nonplanar.

For example, the antibodies of the present invention can usefully be conjugated to filtration media, such as NHS-activated Sepharose or CNBr-activated Sepharose for purposes of immunoaffinity chromatography.

For example, the antibodies of the present invention can usefully be attached to paramagnetic microspheres, typically by biotin-streptavidin interaction, which microspheres can then be used for isolation of cells that express or display the proteins of the present invention. As another example, the antibodies of the present invention can usefully be attached to the surface of a microtiter plate for ELISA.

As noted above, the antibodies of the present invention can be produced in prokaryotic and eukaryotic cells. It is, therefore, another aspect of the present invention to provide cells that express the antibodies of the present invention, including hybridoma cells, B cells, plasma cells, and host cells recombinantly modified to express the antibodies of the present invention.

In yet a further aspect, the present invention provides aptamers evolved to bind specifically to one or more of the proteins and protein fragments of the present invention, to one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention, or the binding of which can be competitively inhibited by one or more of the proteins and protein fragments of the present invention or one or more of the proteins and protein fragments encoded by the isolated nucleic acids of the present invention.

In sum, one of skill in the art, provided with the teachings of this invention, has available a variety of methods which may be used to alter the biological properties of the antibodies of this invention including methods which would increase or decrease the stability or half-life, immunogenicity, toxicity, affinity or yield of a given antibody molecule, or to alter it in any other way that may render it more suitable for a particular application.

Transgenic Animals and Cells

In another aspect, the invention provides transgenic cells and non-human organisms comprising nucleic acid molecules of the invention. In a preferred embodiment, the transgenic cells and non-human organisms comprise a nucleic acid molecule encoding a BSP. In a preferred embodiment, the BSP comprises an amino acid sequence selected from SEQ ID NO: 172 through 295, or a fragment, mutein, homologous protein or allelic variant thereof. In another preferred embodiment, the transgenic cells and non-human organism comprise a BSNA of the invention, preferably a BSNA comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 171, or a part, substantially similar nucleic acid molecule, allelic variant or hybridizing nucleic acid molecule thereof.

In another embodiment, the transgenic cells and non-human organisms have a targeted disruption or replacement of the endogenous orthologue of the human BSG. The transgenic cells can be embryonic stem cells or somatic cells. The transgenic non-human organisms can be chimeric, nonchimeric heterozygotes, and nonchimeric

homozygotes. Methods of producing transgenic animals are well-known in the art. *See, e.g.*, Hogan *et al.*, Manipulating the Mouse Embryo: A Laboratory Manual, 2d ed., Cold Spring Harbor Press (1999); Jackson *et al.*, Mouse Genetics and Transgenics: A Practical Approach, Oxford University Press (2000); and Pinkert, Transgenic Animal Technology: A Laboratory Handbook, Academic Press (1999).

Any technique known in the art may be used to introduce a nucleic acid molecule of the invention into an animal to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection. (*see, e.g.*, Paterson *et al.*, *Appl. Microbiol. Biotechnol.* 40: 691-698 (1994); Carver *et al.*, *Biotechnology* 11: 1263-1270 (1993); Wright *et al.*, *Biotechnology* 9: 830-834 (1991); and U.S. Patent 4,873,191 (1989 retrovirus-mediated gene transfer into germ lines, blastocysts or embryos (*see, e.g.*, Van der Putten *et al.*, *Proc. Natl. Acad. Sci., USA* 82: 6148-6152 (1985)); gene targeting in embryonic stem cells (*see, e.g.*, Thompson *et al.*, *Cell* 56: 313-321 (1989)); electroporation of cells or embryos (*see, e.g.*, Lo, 1983, *Mol. Cell. Biol.* 3: 1803-1814 (1983)); introduction using a gene gun (*see, e.g.*, Ulmer *et al.*, *Science* 259: 1745-49 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm-mediated gene transfer (*see, e.g.*, Lavitrano *et al.*, *Cell* 57: 717-723 (1989)).

Other techniques include, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (*see, e.g.*, Campell *et al.*, *Nature* 380: 64-66 (1996); Wilmut *et al.*, *Nature* 385: 810-813 (1997)). The present invention provides for transgenic animals that carry the transgene (*i.e.*, a nucleic acid molecule of the invention) in all their cells, as well as animals which carry the transgene in some, but not all their cells, i. e., mosaic animals or chimeric animals.

The transgene may be integrated as a single transgene or as multiple copies, such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, *e.g.*, the teaching of Lasko *et al. et al.*, *Proc. Natl. Acad. Sci. USA* 89: 6232-6236 (1992). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to

verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, *in situ* hybridization analysis, and reverse transcriptase-PCR (RT-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Methods for creating a transgenic animal with a disruption of a targeted gene are also well-known in the art. In general, a vector is designed to comprise some nucleotide sequences homologous to the endogenous targeted gene. The vector is introduced into a cell so that it may integrate, via homologous recombination with chromosomal sequences, into the endogenous gene, thereby disrupting the function of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type. *See, e.g.,* Gu *et al., Science* 265: 103-106 (1994). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. *See, e.g.*, Smithies *et al., Nature* 317: 230-234 (1985); Thomas *et al., Cell* 51: 503-512 (1987); Thompson *et al., Cell* 5: 313-321 (1989).

In one embodiment, a mutant, non-functional nucleic acid molecule of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous nucleic acid sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene. *See, e.g.,* Thomas, *supra* and Thompson, *supra*. However this approach can be routinely adapted for use in humans provided the recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (*e.g.*, knockouts) are administered to a patient *in vivo*. Such cells may be obtained from an animal or patient or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (*e.g.*, lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, *e.g.*, by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, *e.g.*, in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, *e.g.*, genetically engineered fibroblasts can be implanted as part of a skin graft;

genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. *See, e.g.,* U.S. Patents 5,399,349 and 5,460,959, each of which is incorporated by reference herein in its entirety.

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well-known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Computer Readable Means

A further aspect of the invention relates to a computer readable means for storing the nucleic acid and amino acid sequences of the instant invention. In a preferred embodiment, the invention provides a computer readable means for storing SEQ ID NO: 1 through 171 and SEQ ID NO: 172 through 295 as described herein, as the complete set of sequences or in any combination. The records of the computer readable means can be accessed for reading and display and for interface with a computer system for the application of programs allowing for the location of data upon a query for data meeting certain criteria, the comparison of sequences, the alignment or ordering of sequences meeting a set of criteria, and the like.

The nucleic acid and amino acid sequences of the invention are particularly useful as components in databases useful for search analyses as well as in sequence analysis algorithms. As used herein, the terms "nucleic acid sequences of the invention" and "amino acid sequences of the invention" mean any detectable chemical or physical characteristic of a polynucleotide or polypeptide of the invention that is or may be reduced to or stored in a computer readable form. These include, without limitation, chromatographic scan data or peak data, photographic data or scan data therefrom, and mass spectrographic data.

This invention provides computer readable media having stored thereon sequences of the invention. A computer readable medium may comprise one or more of the following: a nucleic acid sequence comprising a sequence of a nucleic acid sequence of the invention; an amino acid sequence comprising an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of one or more nucleic acid sequences of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention; a set of nucleic acid sequences wherein at least one of said sequences comprises the sequence of a nucleic acid sequence of the invention; a set of amino acid sequences wherein at least one of said sequences comprises the sequence of an amino acid sequence of the invention; a data set representing a nucleic acid sequence comprising the sequence of a nucleic acid sequence of the invention; a data set representing a nucleic acid sequence encoding an amino acid sequence comprising the sequence of an amino acid sequence of the invention. The computer readable medium can be any composition of matter used to store information or data, including, for example, commercially available floppy disks, tapes, hard drives, compact disks, and video disks.

Also provided by the invention are methods for the analysis of character sequences, particularly genetic sequences. Preferred methods of sequence analysis include, for example, methods of sequence homology analysis, such as identity and similarity analysis, RNA structure analysis, sequence assembly, cladistic analysis, sequence motif analysis, open reading frame determination, nucleic acid base calling, and sequencing chromatogram peak analysis.

A computer-based method is provided for performing nucleic acid sequence identity or similarity identification. This method comprises the steps of providing a nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and comparing said nucleic acid sequence to at least one nucleic acid or amino acid sequence to identify sequence identity or similarity.

A computer-based method is also provided for performing amino acid homology identification, said method comprising the steps of: providing an amino acid sequence comprising the sequence of an amino acid of the invention in a computer readable

medium; and comparing said an amino acid sequence to at least one nucleic acid or an amino acid sequence to identify homology.

A computer-based method is still further provided for assembly of overlapping nucleic acid sequences into a single nucleic acid sequence, said method comprising the steps of: providing a first nucleic acid sequence comprising the sequence of a nucleic acid of the invention in a computer readable medium; and screening for at least one overlapping region between said first nucleic acid sequence and a second nucleic acid sequence.

Diagnostic Methods for Breast Cancer

The present invention also relates to quantitative and qualitative diagnostic assays and methods for detecting, diagnosing, monitoring, staging and predicting cancers by comparing expression of a BSNA or a BSP in a human patient that has or may have breast cancer, or who is at risk of developing breast cancer, with the expression of a BSNA or a BSP in a normal human control. For purposes of the present invention, "expression of a BSNA" or "BSNA expression" means the quantity of BSG mRNA that can be measured by any method known in the art or the level of transcription that can be measured by any method known in the art in a cell, tissue, organ or whole patient. Similarly, the term "expression of a BSP" or "BSP expression" means the amount of BSP that can be measured by any method known in the art or the level of translation of a BSG BSNA that can be measured by any method known in the art.

The present invention provides methods for diagnosing breast cancer in a patient, in particular squamous cell carcinoma, by analyzing for changes in levels of BSNA or BSP in cells, tissues, organs or bodily fluids compared with levels of BSNA or BSP in cells, tissues, organs or bodily fluids of preferably the same type from a normal human control, wherein an increase, or decrease in certain cases, in levels of a BSNA or BSP in the patient versus the normal human control is associated with the presence of breast cancer or with a predilection to the disease. In another preferred embodiment, the present invention provides methods for diagnosing breast cancer in a patient by analyzing changes in the structure of the mRNA of a BSG compared to the mRNA from a normal control. These changes include, without limitation, aberrant splicing, alterations in polyadenylation and/or alterations in 5' nucleotide capping. In yet another preferred embodiment, the present invention provides methods for diagnosing breast cancer in a patient by analyzing changes in a BSP compared to a BSP from a normal control. These

changes include, *e.g.*, alterations in glycosylation and/or phosphorylation of the BSP or subcellular BSP localization.

In a preferred embodiment, the expression of a BSNA is measured by determining the amount of an mRNA that encodes an amino acid sequence selected from SEQ ID NO: 172 through 295, a homolog, an allelic variant, or a fragment thereof. In a more preferred embodiment, the BSNA expression that is measured is the level of expression of a BSNA mRNA selected from SEQ ID NO: 1 through 171, or a hybridizing nucleic acid, homologous nucleic acid or allelic variant thereof, or a part of any of these nucleic acids. BSNA expression may be measured by any method known in the art, such as those described *supra*, including measuring mRNA expression by Northern blot, quantitative or qualitative reverse transcriptase PCR (RT-PCR), microarray, dot or slot blots or *in situ* hybridization. *See, e.g.*, Ausubel (1992), *supra*; Ausubel (1999), *supra*; Sambrook (1989), *supra*; and Sambrook (2001), *supra*. BSNA transcription may be measured by any method known in the art including using a reporter gene hooked up to the promoter of a BSG of interest or doing nuclear run-off assays. Alterations in mRNA structure, *e.g.*, aberrant splicing variants, may be determined by any method known in the art, including, RT-PCR followed by sequencing or restriction analysis. As necessary, BSNA expression may be compared to a known control, such as normal breast nucleic acid, to detect a change in expression.

In another preferred embodiment, the expression of a BSP is measured by determining the level of a BSP having an amino acid sequence selected from the group consisting of SEQ ID NO: 172 through 295, a homolog, an allelic variant, or a fragment thereof. Such levels are preferably determined in at least one of cells, tissues, organs and/or bodily fluids, including determination of normal and abnormal levels. Thus, for instance, a diagnostic assay in accordance with the invention for diagnosing over- or underexpression of BSNA or BSP compared to normal control bodily fluids, cells, or tissue samples may be used to diagnose the presence of breast cancer. The expression level of a BSP may be determined by any method known in the art, such as those described *supra*. In a preferred embodiment, the BSP expression level may be determined by radioimmunoassays, competitive-binding assays, ELISA, Western blot, FACS, immunohistochemistry, immunoprecipitation, proteomic approaches: two-dimensional gel electrophoresis (2D electrophoresis) and non-gel-based approaches such as mass spectrometry or protein interaction profiling. *See*, e.g, Harlow (1999), *supra*; Ausubel (1992), *supra*; and Ausubel (1999), *supra*. Alterations in the BSP

structure may be determined by any method known in the art, including, *e.g.*, using antibodies that specifically recognize phosphoserine, phosphothreonine or phosphotyrosine residues, two-dimensional polyacrylamide gel electrophoresis (2D PAGE) and/or chemical analysis of amino acid residues of the protein. *Id.*

In a preferred embodiment, a radioimmunoassay (RIA) or an ELISA is used. An antibody specific to a BSP is prepared if one is not already available. In a preferred embodiment, the antibody is a monoclonal antibody. The anti-BSP antibody is bound to a solid support and any free protein binding sites on the solid support are blocked with a protein such as bovine serum albumin. A sample of interest is incubated with the antibody on the solid support under conditions in which the BSP will bind to the anti-BSP antibody. The sample is removed, the solid support is washed to remove unbound material, and an anti-BSP antibody that is linked to a detectable reagent (a radioactive substance for RIA and an enzyme for ELISA) is added to the solid support and incubated under conditions in which binding of the BSP to the labeled antibody will occur. After binding, the unbound labeled antibody is removed by washing. For an ELISA, one or more substrates are added to produce a colored reaction product that is based upon the amount of a BSP in the sample. For an RIA, the solid support is counted for radioactive decay signals by any method known in the art. Quantitative results for both RIA and ELISA typically are obtained by reference to a standard curve.

Other methods to measure BSP levels are known in the art. For instance, a competition assay may be employed wherein an anti-BSP antibody is attached to a solid support and an allocated amount of a labeled BSP and a sample of interest are incubated with the solid support. The amount of labeled BSP detected which is attached to the solid support can be correlated to the quantity of a BSP in the sample.

Of the proteomic approaches, 2D PAGE is a well-known technique. Isolation of individual proteins from a sample such as serum is accomplished using sequential separation of proteins by isoelectric point and molecular weight. Typically, polypeptides are first separated by isoelectric point (the first dimension) and then separated by size using an electric current (the second dimension). In general, the second dimension is perpendicular to the first dimension. Because no two proteins with different sequences are identical on the basis of both size and charge, the result of 2D PAGE is a roughly square gel in which each protein occupies a unique spot. Analysis of the spots with chemical or antibody probes, or subsequent protein microsequencing can reveal the relative abundance of a given protein and the identity of the proteins in the sample.

Expression levels of a BSNA can be determined by any method known in the art, including PCR and other nucleic acid methods, such as ligase chain reaction (LCR) and nucleic acid sequence based amplification (NASBA), can be used to detect malignant cells for diagnosis and monitoring of various malignancies. For example, reverse-transcriptase PCR (RT-PCR) is a powerful technique which can be used to detect the presence of a specific mRNA population in a complex mixture of thousands of other mRNA species. In RT-PCR, an mRNA species is first reverse transcribed to complementary DNA (cDNA) with use of the enzyme reverse transcriptase; the cDNA is then amplified as in a standard PCR reaction.

Hybridization to specific DNA molecules (*e.g.*, oligonucleotides) arrayed on a solid support can be used to both detect the expression of and quantitate the level of expression of one or more BSNAs of interest. In this approach, all or a portion of one or more BSNAs is fixed to a substrate. A sample of interest, which may comprise RNA, *e.g.*, total RNA or polyA-selected mRNA, or a complementary DNA (cDNA) copy of the RNA is incubated with the solid support under conditions in which hybridization will occur between the DNA on the solid support and the nucleic acid molecules in the sample of interest. Hybridization between the substrate-bound DNA and the nucleic acid molecules in the sample can be detected and quantitated by several means, including, without limitation, radioactive labeling or fluorescent labeling of the nucleic acid molecule or a secondary molecule designed to detect the hybrid.

The above tests can be carried out on samples derived from a variety of cells, bodily fluids and/or tissue extracts such as homogenates or solubilized tissue obtained from a patient. Tissue extracts are obtained routinely from tissue biopsy and autopsy material. Bodily fluids useful in the present invention include blood, urine, saliva or any other bodily secretion or derivative thereof. By blood it is meant to include whole blood, plasma, serum or any derivative of blood. In a preferred embodiment, the specimen tested for expression of BSNA or BSP includes, without limitation, breast tissue, fluid obtained by bronchial alveolar lavage (BAL), sputum, breast cells grown in cell culture, blood, serum, lymph node tissue and lymphatic fluid. In another preferred embodiment, especially when metastasis of a primary breast cancer is known or suspected, specimens include, without limitation, tissues from brain, bone, bone marrow, liver, adrenal glands and colon. In general, the tissues may be sampled by biopsy, including, without limitation, needle biopsy, *e.g.*, transthoracic needle aspiration, cervical mediatinoscopy, endoscopic lymph node biopsy, video-assisted thoracoscopy, exploratory thoracotomy,

bone marrow biopsy and bone marrow aspiration. *See* Scott, *supra* and Franklin, pp. 529-570, in Kane, *supra*. For early and inexpensive detection, assaying for changes in BSNAs or BSPs in cells in sputum samples may be particularly useful. Methods of obtaining and analyzing sputum samples is disclosed in Franklin, *supra*.

All the methods of the present invention may optionally include determining the expression levels of one or more other cancer markers in addition to determining the expression level of a BSNA or BSP. In many cases, the use of another cancer marker will decrease the likelihood of false positives or false negatives. In one embodiment, the one or more other cancer markers include other BSNA or BSPs as disclosed herein. Other cancer markers useful in the present invention will depend on the cancer being tested and are known to those of skill in the art. In a preferred embodiment, at least one other cancer marker in addition to a particular BSNA or BSP is measured. In a more preferred embodiment, at least two other additional cancer markers are used. In an even more preferred embodiment, at least three, more preferably at least five, even more preferably at least ten additional cancer markers are used.

Diagnosing

In one aspect, the invention provides a method for determining the expression levels and/or structural alterations of one or more BSNAs and/or BSPs in a sample from a patient suspected of having breast cancer. In general, the method comprises the steps of obtaining the sample from the patient, determining the expression level or structural alterations of a BSNA and/or BSP and then ascertaining whether the patient has breast cancer from the expression level of the BSNA or BSP. In general, if high expression relative to a control of a BSNA or BSP is indicative of breast cancer, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of breast cancer, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

The present invention also provides a method of determining whether breast cancer has metastasized in a patient. One may identify whether the breast cancer has metastasized by measuring the expression levels and/or structural alterations of one or more BSNAs and/or BSPs in a variety of tissues. The presence of a BSNA or BSP in a certain tissue at levels higher than that of corresponding noncancerous tissue (*e.g.*, the same tissue from another individual) is indicative of metastasis if high level expression of a BSNA or BSP is associated with breast cancer. Similarly, the presence of a BSNA or BSP in a tissue at levels lower than that of corresponding noncancerous tissue is indicative of metastasis if low level expression of a BSNA or BSP is associated with breast cancer. Further, the presence of a structurally altered BSNA or BSP that is associated with breast cancer is also indicative of metastasis.

In general, if high expression relative to a control of a BSNA or BSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the BSNA or BSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of metastasis, an assay for metastasis is considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control.

The BSNA or BSP of this invention may be used as element in an array or a multi-analyte test to recognize expression patterns associated with breast cancers or other breast related disorders. In addition, the sequences of either the nucleic acids or proteins may be used as elements in a computer program for pattern recognition of breast disorders.

Staging

The invention also provides a method of staging breast cancer in a human patient. The method comprises identifying a human patient having breast cancer and analyzing cells, tissues or bodily fluids from such human patient for expression levels and/or structural alterations of one or more BSNAs or BSPs. First, one or more tumors from a variety of patients are staged according to procedures well-known in the art, and the expression level of one or more BSNAs or BSPs is determined for each stage to obtain a

standard expression level for each BSNA and BSP. Then, the BSNA or BSP expression levels are determined in a biological sample from a patient whose stage of cancer is not known. The BSNA or BSP expression levels from the patient are then compared to the standard expression level. By comparing the expression level of the BSNAs and BSPs from the patient to the standard expression levels, one may determine the stage of the tumor. The same procedure may be followed using structural alterations of a BSNA or BSP to determine the stage of a breast cancer.

Monitoring

Further provided is a method of monitoring breast cancer in a human patient. One may monitor a human patient to determine whether there has been metastasis and, if there has been, when metastasis began to occur. One may also monitor a human patient to determine whether a preneoplastic lesion has become cancerous. One may also monitor a human patient to determine whether a therapy, *e.g.*, chemotherapy, radiotherapy or surgery, has decreased or eliminated the breast cancer. The method comprises identifying a human patient that one wants to monitor for breast cancer, periodically analyzing cells, tissues or bodily fluids from such human patient for expression levels of one or more BSNAs or BSPs, and comparing the BSNA or BSP levels over time to those BSNA or BSP expression levels obtained previously. Patients may also be monitored by measuring one or more structural alterations in a BSNA or BSP that are associated with breast cancer.

If increased expression of a BSNA or BSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an increase in the expression level of a BSNA or BSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. One having ordinary skill in the art would recognize that if this were the case, then a decreased expression level would be indicative of no metastasis, effective therapy or failure to progress to a neoplastic lesion. If decreased expression of a BSNA or BSP is associated with metastasis, treatment failure, or conversion of a preneoplastic lesion to a cancerous lesion, then detecting an decrease in the expression level of a BSNA or BSP indicates that the tumor is metastasizing, that treatment has failed or that the lesion is cancerous, respectively. In a preferred embodiment, the levels of BSNAs or BSPs are determined from the same cell type, tissue or bodily fluid as prior patient samples.

Monitoring a patient for onset of breast cancer metastasis is periodic and preferably is done on a quarterly basis, but may be done more or less frequently.

The methods described herein can further be utilized as prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with increased or decreased expression levels of a BSNA and/or BSP. The present invention provides a method in which a test sample is obtained from a human patient and one or more BSNAs and/or BSPs are detected. The presence of higher (or lower) BSNA or BSP levels as compared to normal human controls is diagnostic for the human patient being at risk for developing cancer, particularly breast cancer. The effectiveness of therapeutic agents to decrease (or increase) expression or activity of one or more BSNAs and/or BSPs of the invention can also be monitored by analyzing levels of expression of the BSNAs and/or BSPs in a human patient in clinical trials or in *in vitro* screening assays such as in human cells. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the human patient or cells, as the case may be, to the agent being tested.

Detection of Genetic Lesions or Mutations

The methods of the present invention can also be used to detect genetic lesions or mutations in a BSG, thereby determining if a human with the genetic lesion is susceptible to developing breast cancer or to determine what genetic lesions are responsible, or are partly responsible, for a person's existing breast cancer. Genetic lesions can be detected, for example, by ascertaining the existence of a deletion, insertion and/or substitution of one or more nucleotides from the BSGs of this invention, a chromosomal rearrangement of BSG, an aberrant modification of BSG (such as of the methylation pattern of the genomic DNA), or allelic loss of a BSG. Methods to detect such lesions in the BSG of this invention are known to those having ordinary skill in the art following the teachings of the specification.

<u>Methods of Detecting Noncancerous Breast Diseases</u>

The invention also provides a method for determining the expression levels and/or structural alterations of one or more BSNAs and/or BSPs in a sample from a patient suspected of having or known to have a noncancerous breast disease. In general, the method comprises the steps of obtaining a sample from the patient, determining the expression level or structural alterations of a BSNA and/or BSP, comparing the

expression level or structural alteration of the BSNA or BSP to a normal breast control, and then ascertaining whether the patient has a noncancerous breast disease. In general, if high expression relative to a control of a BSNA or BSP is indicative of a particular noncancerous breast disease, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times higher, and more preferably are at least five times higher, even more preferably at least ten times higher, than in preferably the same cells, tissues or bodily fluid of a normal human control. In contrast, if low expression relative to a control of a BSNA or BSP is indicative of a noncancerous breast disease, a diagnostic assay is considered positive if the level of expression of the BSNA or BSP is at least two times lower, more preferably are at least five times lower, even more preferably at least ten times lower than in preferably the same cells, tissues or bodily fluid of a normal human control. The normal human control may be from a different patient or from uninvolved tissue of the same patient.

One having ordinary skill in the art may determine whether a BSNA and/or BSP is associated with a particular noncancerous breast disease by obtaining breast tissue from a patient having a noncancerous breast disease of interest and determining which BSNAs and/or BSPs are expressed in the tissue at either a higher or a lower level than in normal breast tissue. In another embodiment, one may determine whether a BSNA or BSP exhibits structural alterations in a particular noncancerous breast disease state by obtaining breast tissue from a patient having a noncancerous breast disease of interest and determining the structural alterations in one or more BSNAs and/or BSPs relative to normal breast tissue.

Methods for Identifying Breast Tissue

In another aspect, the invention provides methods for identifying breast tissue. These methods are particularly useful in, *e.g.*, forensic science, breast cell differentiation and development, and in tissue engineering.

In one embodiment, the invention provides a method for determining whether a sample is breast tissue or has breast tissue-like characteristics. The method comprises the steps of providing a sample suspected of comprising breast tissue or having breast tissue-like characteristics, determining whether the sample expresses one or more BSNAs and/or BSPs, and, if the sample expresses one or more BSNAs and/or BSPs, concluding that the sample comprises breast tissue. In a preferred embodiment, the BSNA encodes a polypeptide having an amino acid sequence selected from SEQ ID NO: 172 through 295,

or a homolog, allelic variant or fragment thereof. In a more preferred embodiment, the BSNA has a nucleotide sequence selected from SEQ ID NO: 1 through 171, or a hybridizing nucleic acid, an allelic variant or a part thereof. Determining whether a sample expresses a BSNA can be accomplished by any method known in the art. Preferred methods include hybridization to microarrays, Northern blot hybridization, and quantitative or qualitative RT-PCR. In another preferred embodiment, the method can be practiced by determining whether a BSP is expressed. Determining whether a sample expresses a BSP can be accomplished by any method known in the art. Preferred methods include Western blot, ELISA, RIA and 2D PAGE. In one embodiment, the BSP has an amino acid sequence selected from SEQ ID NO: 172 through 295, or a homolog, allelic variant or fragment thereof. In another preferred embodiment, the expression of at least two BSNAs and/or BSPs is determined. In a more preferred embodiment, the expression of at least three, more preferably four and even more preferably five BSNAs and/or BSPs are determined.

In one embodiment, the method can be used to determine whether an unknown tissue is breast tissue. This is particularly useful in forensic science, in which small, damaged pieces of tissues that are not identifiable by microscopic or other means are recovered from a crime or accident scene. In another embodiment, the method can be used to determine whether a tissue is differentiating or developing into breast tissue. This is important in monitoring the effects of the addition of various agents to cell or tissue culture, *e.g.*, in producing new breast tissue by tissue engineering. These agents include, *e.g.*, growth and differentiation factors, extracellular matrix proteins and culture medium. Other factors that may be measured for effects on tissue development and differentiation include gene transfer into the cells or tissues, alterations in pH, aqueous:air interface and various other culture conditions.

Methods for Producing and Modifying Breast Tissue

In another aspect, the invention provides methods for producing engineered breast tissue or cells. In one embodiment, the method comprises the steps of providing cells, introducing a BSNA or a BSG into the cells, and growing the cells under conditions in which they exhibit one or more properties of breast tissue cells. In a preferred embodiment, the cells are pluripotent. As is well-known in the art, normal breast tissue comprises a large number of different cell types. Thus, in one embodiment, the engineered breast tissue or cells comprises one of these cell types. In another

embodiment, the engineered breast tissue or cells comprises more than one breast cell type. Further, the culture conditions of the cells or tissue may require manipulation in order to achieve full differentiation and development of the breast cell tissue. Methods for manipulating culture conditions are well-known in the art.

Nucleic acid molecules encoding one or more BSPs are introduced into cells, preferably pluripotent cells. In a preferred embodiment, the nucleic acid molecules encode BSPs having amino acid sequences selected from SEQ ID NO: 172 through 295, or homologous proteins, analogs, allelic variants or fragments thereof. In a more preferred embodiment, the nucleic acid molecules have a nucleotide sequence selected from SEQ ID NO: 1 through 171, or hybridizing nucleic acids, allelic variants or parts thereof. In another highly preferred embodiment, a BSG is introduced into the cells. Expression vectors and methods of introducing nucleic acid molecules into cells are well-known in the art and are described in detail, *supra.*

Artificial breast tissue may be used to treat patients who have lost some or all of their breast function.

Pharmaceutical Compositions

In another aspect, the invention provides pharmaceutical compositions comprising the nucleic acid molecules, polypeptides, antibodies, antibody derivatives, antibody fragments, agonists, antagonists, and inhibitors of the present invention. In a preferred embodiment, the pharmaceutical composition comprises a BSNA or part thereof. In a more preferred embodiment, the BSNA has a nucleotide sequence selected from the group consisting of SEQ ID NO: 1 through 171, a nucleic acid that hybridizes thereto, an allelic variant thereof, or a nucleic acid that has substantial sequence identity thereto. In another preferred embodiment, the pharmaceutical composition comprises a BSP or fragment thereof. In a more preferred embodiment, the BSP having an amino acid sequence that is selected from the group consisting of SEQ ID NO: 172 through 295, a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof. In another preferred embodiment, the pharmaceutical composition comprises an anti-BSP antibody, preferably an antibody that specifically binds to a BSP having an amino acid that is selected from the group consisting of SEQ ID NO: 172 through 295, or an antibody that binds to a polypeptide that is homologous thereto, a fusion protein comprising all or a portion of the polypeptide, or an analog or derivative thereof.

Such a composition typically contains from about 0.1 to 90% by weight of a therapeutic agent of the invention formulated in and/or with a pharmaceutically acceptable carrier or excipient.

Pharmaceutical formulation is a well-established art, and is further described in Gennaro (ed.), Remington: The Science and Practice of Pharmacy, 20th ed., Lippincott, Williams & Wilkins (2000); Ansel *et al.*, Pharmaceutical Dosage Forms and Drug Delivery Systems, 7th ed., Lippincott Williams & Wilkins (1999); and Kibbe (ed.), Handbook of Pharmaceutical Excipients American Pharmaceutical Association, 3rd ed. (2000), the disclosures of which are incorporated herein by reference in their entireties, and thus need not be described in detail herein.

Briefly, formulation of the pharmaceutical compositions of the present invention will depend upon the route chosen for administration. The pharmaceutical compositions utilized in this invention can be administered by various routes including both enteral and parenteral routes, including oral, intravenous, intramuscular, subcutaneous, inhalation, topical, sublingual, rectal, intra-arterial, intramedullary, intrathecal, intraventricular, transmucosal, transdermal, intranasal, intraperitoneal, intrapulmonary, and intrauterine.

Oral dosage forms can be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Solid formulations of the compositions for oral administration can contain suitable carriers or excipients, such as carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, or microcrystalline cellulose; gums including arabic and tragacanth; proteins such as gelatin and collagen; inorganics, such as kaolin, calcium carbonate, dicalcium phosphate, sodium chloride; and other agents such as acacia and alginic acid.

Agents that facilitate disintegration and/or solubilization can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate, microcrystalline cellulose, corn starch, sodium starch glycolate, and alginic acid.

Tablet binders that can be used include acacia, methylcellulose, sodium carboxymethylcellulose, polyvinylpyrrolidone (Povidone™), hydroxypropyl methylcellulose, sucrose, starch and ethylcellulose.

Lubricants that can be used include magnesium stearates, stearic acid, silicone fluid, talc, waxes, oils, and colloidal silica.

Fillers, agents that facilitate disintegration and/or solubilization, tablet binders and lubricants, including the aforementioned, can be used singly or in combination.

Solid oral dosage forms need not be uniform throughout. For example, dragee cores can be used in conjunction with suitable coatings, such as concentrated sugar solutions, which can also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures.

Oral dosage forms of the present invention include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Additionally, dyestuffs or pigments can be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, *i.e.*, dosage.

Liquid formulations of the pharmaceutical compositions for oral (enteral) administration are prepared in water or other aqueous vehicles and can contain various suspending agents such as methylcellulose, alginates, tragacanth, pectin, kelgin, carrageenan, acacia, polyvinylpyrrolidone, and polyvinyl alcohol. The liquid formulations can also include solutions, emulsions, syrups and elixirs containing, together with the active compound(s), wetting agents, sweeteners, and coloring and flavoring agents.

The pharmaceutical compositions of the present invention can also be formulated for parenteral administration. Formulations for parenteral administration can be in the form of aqueous or non-aqueous isotonic sterile injection solutions or suspensions.

For intravenous injection, water soluble versions of the compounds of the present invention are formulated in, or if provided as a lyophilate, mixed with, a physiologically acceptable fluid vehicle, such as 5% dextrose ("D5"), physiologically buffered saline, 0.9% saline, Hanks' solution, or Ringer's solution. Intravenous formulations may include carriers, excipients or stabilizers including, without limitation, calcium, human serum albumin, citrate, acetate, calcium chloride, carbonate, and other salts.

Intramuscular preparations, *e.g.* a sterile formulation of a suitable soluble salt form of the compounds of the present invention, can be dissolved and administered in a pharmaceutical excipient such as Water-for-Injection, 0.9% saline, or 5% glucose solution. Alternatively, a suitable insoluble form of the compound can be prepared and administered as a suspension in an aqueous base or a pharmaceutically acceptable oil base, such as an ester of a long chain fatty acid (*e.g.*, ethyl oleate), fatty oils such as sesame oil, triglycerides, or liposomes.

Parenteral formulations of the compositions can contain various carriers such as vegetable oils, dimethylacetamide, dimethylformamide, ethyl lactate, ethyl carbonate, isopropyl myristate, ethanol, polyols (glycerol, propylene glycol, liquid polyethylene glycol, and the like).

Aqueous injection suspensions can also contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Non-lipid polycationic amino polymers can also be used for delivery. Optionally, the suspension can also contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

Pharmaceutical compositions of the present invention can also be formulated to permit injectable, long-term, deposition. Injectable depot forms may be made by forming microencapsulated matrices of the compound in biodegradable polymers such as polylactide-polyglycolide. Depending upon the ratio of drug to polymer and the nature of the particular polymer employed, the rate of drug release can be controlled. Examples of other biodegradable polymers include poly(orthoesters) and poly(anhydrides). Depot injectable formulations are also prepared by entrapping the drug in microemulsions that are compatible with body tissues.

The pharmaceutical compositions of the present invention can be administered topically.

For topical use the compounds of the present invention can also be prepared in suitable forms to be applied to the skin, or mucus membranes of the nose and throat, and can take the form of lotions, creams, ointments, liquid sprays or inhalants, drops, tinctures, lozenges, or throat paints. Such topical formulations further can include chemical compounds such as dimethylsulfoxide (DMSO) to facilitate surface penetration of the active ingredient. In other transdermal formulations, typically in patch-delivered formulations, the pharmaceutically active compound is formulated with one or more skin penetrants, such as 2-N-methyl-pyrrolidone (NMP) or Azone. A topical semi-solid

ointment formulation typically contains a concentration of the active ingredient from about 1 to 20%, *e.g.*, 5 to 10%, in a carrier such as a pharmaceutical cream base.

For application to the eyes or ears, the compounds of the present invention can be presented in liquid or semi-liquid form formulated in hydrophobic or hydrophilic bases as ointments, creams, lotions, paints or powders.

For rectal administration the compounds of the present invention can be administered in the form of suppositories admixed with conventional carriers such as cocoa butter, wax or other glyceride.

Inhalation formulations can also readily be formulated. For inhalation, various powder and liquid formulations can be prepared. For aerosol preparations, a sterile formulation of the compound or salt form of the compound may be used in inhalers, such as metered dose inhalers, and nebulizers. Aerosolized forms may be especially useful for treating respiratory disorders.

Alternatively, the compounds of the present invention can be in powder form for reconstitution in the appropriate pharmaceutically acceptable carrier at the time of delivery.

The pharmaceutically active compound in the pharmaceutical compositions of the present invention can be provided as the salt of a variety of acids, including but not limited to hydrochloric, sulfuric, acetic, lactic, tartaric, malic, and succinic acid. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms.

After pharmaceutical compositions have been prepared, they are packaged in an appropriate container and labeled for treatment of an indicated condition.

The active compound will be present in an amount effective to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

A "therapeutically effective dose" refers to that amount of active ingredient, for example BSP polypeptide, fusion protein, or fragments thereof, antibodies specific for BSP, agonists, antagonists or inhibitors of BSP, which ameliorates the signs or symptoms of the disease or prevents progression thereof; as would be understood in the medical arts, cure, although desired, is not required.

The therapeutically effective dose of the pharmaceutical agents of the present invention can be estimated initially by *in vitro* tests, such as cell culture assays, followed by assay in model animals, usually mice, rats, rabbits, dogs, or pigs. The animal model

can also be used to determine an initial preferred concentration range and route of administration.

For example, the ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population) can be determined in one or more cell culture of animal model systems. The dose ratio of toxic to therapeutic effects is the therapeutic index, which can be expressed as LD50/ED50. Pharmaceutical compositions that exhibit large therapeutic indices are preferred.

The data obtained from cell culture assays and animal studies are used in formulating an initial dosage range for human use, and preferably provide a range of circulating concentrations that includes the ED50 with little or no toxicity. After administration, or between successive administrations, the circulating concentration of active agent varies within this range depending upon pharmacokinetic factors well-known in the art, such as the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors specific to the subject requiring treatment. Factors that can be taken into account by the practitioner include the severity of the disease state, general health of the subject, age, weight, gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Where the therapeutic agent is a protein or antibody of the present invention, the therapeutic protein or antibody agent typically is administered at a daily dosage of 0.01 mg to 30 mg/kg of body weight of the patient (*e.g.*, 1 mg/kg to 5 mg/kg). The pharmaceutical formulation can be administered in multiple doses per day, if desired, to achieve the total desired daily dose.

Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

Conventional methods, known to those of ordinary skill in the art of medicine, can be used to administer the pharmaceutical formulation(s) of the present invention to

the patient. The pharmaceutical compositions of the present invention can be administered alone, or in combination with other therapeutic agents or interventions.

Therapeutic Methods

The present invention further provides methods of treating subjects having defects in a gene of the invention , *e.g.*, in expression, activity, distribution, localization, and/or solubility, which can manifest as a disorder of breast function. As used herein, "treating" includes all medically-acceptable types of therapeutic intervention, including palliation and prophylaxis (prevention) of disease. The term "treating" encompasses any improvement of a disease, including minor improvements. These methods are discussed below.

Gene Therapy and Vaccines

The isolated nucleic acids of the present invention can also be used to drive *in vivo* expression of the polypeptides of the present invention. *In vivo* expression can be driven from a vector, typically a viral vector, often a vector based upon a replication incompetent retrovirus, an adenovirus, or an adeno-associated virus (AAV) , for purpose of gene therapy. *In vivo* expression can also be driven from signals endogenous to the nucleic acid or from a vector, often a plasmid vector, such as pVAX1 (Invitrogen, Carlsbad, CA, USA), for purpose of "naked" nucleic acid vaccination, as further described in U.S. Patents 5,589,466; 5,679,647; 5,804,566; 5,830,877; 5,843,913; 5,880,104; 5,958,891; 5,985,847; 6,017,897; 6,110,898; and 6,204,250, the disclosures of which are incorporated herein by reference in their entireties. For cancer therapy, it is preferred that the vector also be tumor-selective. *See, e.g.*, Doronin *et al.*, *J. Virol.* 75: 3314-24 (2001).

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a nucleic acid of the present invention is administered. The nucleic acid can be delivered in a vector that drives expression of a BSP, fusion protein, or fragment thereof, or without such vector. Nucleic acid compositions that can drive expression of a BSP are administered, for example, to complement a deficiency in the native BSP, or as DNA vaccines. Expression vectors derived from virus, replication deficient retroviruses, adenovirus, adeno-associated (AAV) virus, herpes virus, or vaccinia virus can be used as can plasmids. *See, e.g.*, Cid-Arregui, *supra*. In a preferred embodiment, the nucleic acid

molecule encodes a BSP having the amino acid sequence of SEQ ID NO: 172 through 295, or a fragment, fusion protein, allelic variant or homolog thereof.

In still other therapeutic methods of the present invention, pharmaceutical compositions comprising host cells that express a BSP, fusions, or fragments thereof can be administered. In such cases, the cells are typically autologous, so as to circumvent xenogeneic or allotypic rejection, and are administered to complement defects in BSP production or activity. In a preferred embodiment, the nucleic acid molecules in the cells encode a BSP having the amino acid sequence of SEQ ID NO: 172 through 295, or a fragment, fusion protein, allelic variant or homolog thereof.

Antisense Administration

Antisense nucleic acid compositions, or vectors that drive expression of a BSG antisense nucleic acid, are administered to downregulate transcription and/or translation of a BSG in circumstances in which excessive production, or production of aberrant protein, is the pathophysiologic basis of disease.

Antisense compositions useful in therapy can have a sequence that is complementary to coding or to noncoding regions of a BSG. For example, oligonucleotides derived from the transcription initiation site, *e.g.*, between positions -10 and +10 from the start site, are preferred.

Catalytic antisense compositions, such as ribozymes, that are capable of sequence-specific hybridization to BSG transcripts, are also useful in therapy. *See, e.g.*, Phylactou, *Adv. Drug Deliv. Rev.* 44(2-3): 97-108 (2000); Phylactou *et al.*, *Hum. Mol. Genet.* 7(10): 1649-53 (1998); Rossi, *Ciba Found. Symp.* 209: 195-204 (1997); and Sigurdsson *et al.*, *Trends Biotechnol.* 13(8): 286-9 (1995), the disclosures of which are incorporated herein by reference in their entireties.

Other nucleic acids useful in the therapeutic methods of the present invention are those that are capable of triplex helix formation in or near the BSG genomic locus. Such triplexing oligonucleotides are able to inhibit transcription. *See, e.g.*, Intody *et al.*, *Nucleic Acids Res.* 28(21): 4283-90 (2000); McGuffie *et al.*, *Cancer Res.* 60(14): 3790-9 (2000), the disclosures of which are incorporated herein by reference. Pharmaceutical compositions comprising such triplex forming oligos (TFOs) are administered in circumstances in which excessive production, or production of aberrant protein, is a pathophysiologic basis of disease.

In a preferred embodiment, the antisense molecule is derived from a nucleic acid molecule encoding a BSP, preferably a BSP comprising an amino acid sequence of SEQ ID NO: 172 through 295, or a fragment, allelic variant or homolog thereof. In a more preferred embodiment, the antisense molecule is derived from a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 171, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Polypeptide Administration

In one embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising a BSP, a fusion protein, fragment, analog or derivative thereof is administered to a subject with a clinically-significant BSP defect.

Protein compositions are administered, for example, to complement a deficiency in native BSP. In other embodiments, protein compositions are administered as a vaccine to elicit a humoral and/or cellular immune response to BSP. The immune response can be used to modulate activity of BSP or, depending on the immunogen, to immunize against aberrant or aberrantly expressed forms, such as mutant or inappropriately expressed isoforms. In yet other embodiments, protein fusions having a toxic moiety are administered to ablate cells that aberrantly accumulate BSP.

In a preferred embodiment, the polypeptide is a BSP comprising an amino acid sequence of SEQ ID NO: 172 through 295, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the polypeptide is encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 171, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Antibody, Agonist and Antagonist Administration

In another embodiment of the therapeutic methods of the present invention, a therapeutically effective amount of a pharmaceutical composition comprising an antibody (including fragment or derivative thereof) of the present invention is administered. As is well-known, antibody compositions are administered, for example, to antagonize activity of BSP, or to target therapeutic agents to sites of BSP presence and/or accumulation. In a preferred embodiment, the antibody specifically binds to a BSP comprising an amino acid sequence of SEQ ID NO: 172 through 295, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred

embodiment, the antibody specifically binds to a BSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 171, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

The present invention also provides methods for identifying modulators which bind to a BSP or have a modulatory effect on the expression or activity of a BSP. Modulators which decrease the expression or activity of BSP (antagonists) are believed to be useful in treating breast cancer. Such screening assays are known to those of skill in the art and include, without limitation, cell-based assays and cell-free assays. Small molecules predicted via computer imaging to specifically bind to regions of a BSP can also be designed, synthesized and tested for use in the imaging and treatment of breast cancer. Further, libraries of molecules can be screened for potential anticancer agents by assessing the ability of the molecule to bind to the BSPs identified herein. Molecules identified in the library as being capable of binding to a BSP are key candidates for further evaluation for use in the treatment of breast cancer. In a preferred embodiment, these molecules will downregulate expression and/or activity of a BSP in cells.

In another embodiment of the therapeutic methods of the present invention, a pharmaceutical composition comprising a non-antibody antagonist of BSP is administered. Antagonists of BSP can be produced using methods generally known in the art. In particular, purified BSP can be used to screen libraries of pharmaceutical agents, often combinatorial libraries of small molecules, to identify those that specifically bind and antagonize at least one activity of a BSP.

In other embodiments a pharmaceutical composition comprising an agonist of a BSP is administered. Agonists can be identified using methods analogous to those used to identify antagonists.

In a preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a BSP comprising an amino acid sequence of SEQ ID NO: 172 through 295, or a fusion protein, allelic variant, homolog, analog or derivative thereof. In a more preferred embodiment, the antagonist or agonist specifically binds to and antagonizes or agonizes, respectively, a BSP encoded by a nucleic acid molecule having a nucleotide sequence of SEQ ID NO: 1 through 171, or a part, allelic variant, substantially similar or hybridizing nucleic acid thereof.

Targeting Breast Tissue

The invention also provides a method in which a polypeptide of the invention, or an antibody thereto, is linked to a therapeutic agent such that it can be delivered to the

breast or to specific cells in the breast. In a preferred embodiment, an anti-BSP antibody is linked to a therapeutic agent and is administered to a patient in need of such therapeutic agent. The therapeutic agent may be a toxin, if breast tissue needs to be selectively destroyed. This would be useful for targeting and killing breast cancer cells. In another embodiment, the therapeutic agent may be a growth or differentiation factor, which would be useful for promoting breast cell function.

In another embodiment, an anti-BSP antibody may be linked to an imaging agent that can be detected using, *e.g.*, magnetic resonance imaging, CT or PET. This would be useful for determining and monitoring breast function, identifying breast cancer tumors, and identifying noncancerous breast diseases.

EXAMPLES

Example 1: Gene Expression analysis

BSGs were identified by mRNA subtraction analysis using standard methods. The sequences were extended using GeneBank sequences, Incyte's proprietary database. From the nucleotide sequences, predicted amino acid sequences were prepared. DEX0306_1, DEX0306_2 correspond to SEQ ID NO.1, 2 etc. DEX0157 was the parent sequence found in the mRNA subtractions.

```
DEX0306_1    DEX0157_1    DEX0306_172
DEX0306_2    flex DEX0157_1
DEX0306_3    DEX0157_2    DEX0306_173
DEX0306_4    flex DEX0157_2
DEX0306_5    DEX0157_3    DEX0306_174
DEX0306_6    flex DEX0157_3
DEX0306_7    DEX0157_4    DEX0306_175
DEX0306_8    flex DEX0157_4
DEX0306_9    DEX0157_5    DEX0306_176
DEX0306_10   flex DEX0157_5
DEX0306_11   DEX0157_6    DEX0306_177
DEX0306_12   flex DEX0157_6
DEX0306_13   DEX0157_7    DEX0306_178
DEX0306_14   DEX0157_8    DEX0306_179
DEX0306_15   DEX0157_9    DEX0306_180
DEX0306_16   flex DEX0157_9
DEX0306_17   DEX0157_10   DEX0306_181
DEX0306_18   flex DEX0157_10    DEX0306_182
DEX0306_19   DEX0157_11   DEX0306_183
DEX0306_20   flex DEX0157_11
DEX0306_21   DEX0157_12   DEX0306_184
DEX0306_22   flex DEX0157_12
DEX0306_23   DEX0157_13   DEX0306_185
DEX0306_24   flex DEX0157_13
DEX0306_25   DEX0157_14   DEX0306_186
DEX0306_26   flex DEX0157_14
DEX0306_27   DEX0157_15   DEX0306_187
DEX0306_28   flex DEX0157_15
DEX0306_29   DEX0157_16   DEX0306_188
```

DEX0306_30	DEX0157_17	DEX0306_189
DEX0306_31	flex DEX0157_17	DEX0306_190
DEX0306_32	DEX0157_18	DEX0306_191
DEX0306_33	flex DEX0157_18	
DEX0306_34	DEX0157_19	DEX0306_192
DEX0306_35	DEX0157_20	DEX0306_193
DEX0306_36	flex DEX0157_20	DEX0306_194
DEX0306_37	DEX0157_21	
DEX0306_38	DEX0157_22	DEX0306_195
DEX0306_39	flex DEX0157_22	
DEX0306_40	DEX0157_23	DEX0306_196
DEX0306_41	flex DEX0157_23	
DEX0306_42	DEX0157_24	DEX0306_197
DEX0306_43	DEX0157_25	DEX0306_198
DEX0306_44	flex DEX0157_25	DEX0306_199
DEX0306_45	DEX0157_26	DEX0306_200
DEX0306_46	DEX0157_27	DEX0306_201
DEX0306_47	flex DEX0157_27	
DEX0306_48	DEX0157_28	DEX0306_202
DEX0306_49	flex DEX0157_28	
DEX0306_50	DEX0157_29	DEX0306_203
DEX0306_51	flex DEX0157_29	
DEX0306_52	DEX0157_30	DEX0306_204
DEX0306_53	flex DEX0157_30	DEX0306_205
DEX0306_54	DEX0157_31	DEX0306_206
DEX0306_55	flex DEX0157_31	
DEX0306_56	DEX0157_32	DEX0306_207
DEX0306_57	flex DEX0157_32	
DEX0306_58	DEX0157_33	DEX0306_208
DEX0306_59	flex DEX0157_33	
DEX0306_60	DEX0157_34	
DEX0306_61	flex DEX0157_34	
DEX0306_62	DEX0157_35	DEX0306_209
DEX0306_63	DEX0157_36	DEX0306_210
DEX0306_64	flex DEX0157_36	
DEX0306_65	DEX0157_37	DEX0306_211
DEX0306_66	flex DEX0157_37	DEX0306_212
DEX0306_67	DEX0157_38	DEX0306_213
DEX0306_68	DEX0157_39	DEX0306_214
DEX0306_69	flex DEX0157_39	DEX0306_215
DEX0306_70	DEX0157_40	DEX0306_216
DEX0306_71	flex DEX0157_40	DEX0306_217
DEX0306_72	DEX0157_41	DEX0306_218
DEX0306_73	flex DEX0157_41	DEX0306_219
DEX0306_74	DEX0157_42	DEX0306_220
DEX0306_75	flex DEX0157_42	
DEX0306_76	DEX0157_43	DEX0306_221
DEX0306_77	flex DEX0157_43	
DEX0306_78	DEX0157_44	DEX0306_222
DEX0306_79	flex DEX0157_44	
DEX0306_80	DEX0157_45	DEX0306_223
DEX0306_81	flex DEX0157_45	DEX0306_224
DEX0306_82	DEX0157_46	DEX0306_225
DEX0306_83	DEX0157_47	DEX0306_226
DEX0306_84	DEX0157_48	DEX0306_227
DEX0306_85	DEX0157_49	DEX0306_228
DEX0306_86	flex DEX0157_49	DEX0306_229
DEX0306_87	DEX0157_50	DEX0306_230
DEX0306_88	flex DEX0157_50	DEX0306_231
DEX0306_89	DEX0157_51	DEX0306_232
DEX0306_90	flex DEX0157_51	
DEX0306_91	DEX0157_52	DEX0306_233

DEX0306_92 DEX0157_53 DEX0306_234
DEX0306_93 flex DEX0157_53 DEX0306_235
DEX0306_94 DEX0157_54 DEX0306_236
DEX0306_95 flex DEX0157_54
DEX0306_96 DEX0157_55 DEX0306_237
DEX0306_97 DEX0157_56 DEX0306_238
DEX0306_98 flex DEX0157_56 DEX0306_239
DEX0306_99 DEX0157_57 DEX0306_240
DEX0306_100 DEX0157_58 DEX0306_241
DEX0306_101 flex DEX0157_58
DEX0306_102 DEX0157_60 DEX0306_242
DEX0306_103 flex DEX0157_60 DEX0306_243
DEX0306_104 DEX0157_61 DEX0306_244
DEX0306_105 flex DEX0157_61 DEX0306_245
DEX0306_106 DEX0157_62 DEX0306_246
DEX0306_107 flex DEX0157_62 DEX0306_247
DEX0306_108 DEX0157_63 DEX0306_248
DEX0306_109 flex DEX0157_63
DEX0306_110 DEX0157_64 DEX0306_249
DEX0306_111 flex DEX0157_64 DEX0306_250
DEX0306_112 DEX0157_65 DEX0306_251
DEX0306_113 DEX0157_66 DEX0306_252
DEX0306_114 DEX0157_67 DEX0306_253
DEX0306_115 DEX0157_68 DEX0306_254
DEX0306_116 flex DEX0157_68 DEX0306_255
DEX0306_117 DEX0157_69 DEX0306_256
DEX0306_118 flex DEX0157_69 DEX0306_257
DEX0306_119 DEX0157_70 DEX0306_258
DEX0306_120 flex DEX0157_70
DEX0306_121 DEX0157_71 DEX0306_259
DEX0306_122 flex DEX0157_71
DEX0306_123 DEX0157_72 DEX0306_260
DEX0306_124 flex DEX0157_72 DEX0306_261
DEX0306_125 DEX0157_73 DEX0306_262
DEX0306_126 flex DEX0157_73 DEX0306_263
DEX0306_127 DEX0157_74 DEX0306_264
DEX0306_128 flex DEX0157_74
DEX0306_129 DEX0157_75 DEX0306_265
DEX0306_130 DEX0157_76 DEX0306_266
DEX0306_131 flex DEX0157_76 DEX0306_267
DEX0306_132 DEX0157_77 DEX0306_268
DEX0306_133 flex DEX0157_77
DEX0306_134 DEX0157_78 DEX0306_269
DEX0306_135 flex DEX0157_78 DEX0306_270
DEX0306_136 DEX0157_79 DEX0306_271
DEX0306_137 flex DEX0157_79 DEX0306_272
DEX0306_138 DEX0157_80 DEX0306_273
DEX0306_139 DEX0157_81 DEX0306_274
DEX0306_140 flex DEX0157_81 DEX0306_275
DEX0306_141 DEX0157_82 DEX0306_276
DEX0306_142 flex DEX0157_82
DEX0306_143 DEX0157_83 DEX0306_277
DEX0306_144 flex DEX0157_83
DEX0306_145 DEX0157_85 DEX0306_278
DEX0306_146 flex DEX0157_85
DEX0306_147 DEX0157_86 DEX0306_279
DEX0306_148 flex DEX0157_86 DEX0306_280
DEX0306_149 DEX0157_87 DEX0306_281
DEX0306_150 flex DEX0157_87
DEX0306_151 DEX0157_88 DEX0306_282
DEX0306_152 flex DEX0157_88
DEX0306_153 DEX0157_89 DEX0306_283

DEX0306_154 flex DEX0157_89
DEX0306_155 DEX0157_90 DEX0306_284
DEX0306_156 flex DEX0157_90 DEX0306_285
DEX0306_157 DEX0157_93 DEX0306_286
DEX0306_158 DEX0157_94 DEX0306_287
DEX0306_159 flex DEX0157_94
DEX0306_160 DEX0157_95 DEX0306_288
DEX0306_161 flex DEX0157_95
DEX0306_162 DEX0157_96 DEX0306_289
DEX0306_163 DEX0157_97 DEX0306_290
DEX0306_164 flex DEX0157_97
DEX0306_165 DEX0157_98 DEX0306_291
DEX0306_166 DEX0157_99 DEX0306_292
DEX0306_167 DEX0157_100 DEX0306_293
DEX0306_168 flex DEX0157_100
DEX0306_169 DEX0157_101 DEX0306_294
DEX0306_170 DEX0157_102 DEX0306_295
DEX0306_171 flex DEX0157_102

Example 2: Relative Quantitation of Gene Expression

Real-Time quantitative PCR with fluorescent Taqman probes is a quantitation detection system utilizing the 5'- 3' nuclease activity of Taq DNA polymerase. The method uses an internal fluorescent oligonucleotide probe (Taqman) labeled with a 5' reporter dye and a downstream, 3' quencher dye. During PCR, the 5'-3' nuclease activity of Taq DNA polymerase releases the reporter, whose fluorescence can then be detected by the laser detector of the Model 7700 Sequence Detection System (PE Applied Biosystems, Foster City, CA, USA). Amplification of an endogenous control is used to standardize the amount of sample RNA added to the reaction and normalize for Reverse Transcriptase (RT) efficiency. Either cyclophilin, glyceraldehyde-3-phosphate dehydrogenase (GAPDH), ATPase, or 18S ribosomal RNA (rRNA) is used as this endogenous control. To calculate relative quantitation between all the samples studied, the target RNA levels for one sample were used as the basis for comparative results (calibrator). Quantitation relative to the "calibrator" can be obtained using the standard curve method or the comparative method (User Bulletin #2: ABI PRISM 7700 Sequence Detection System).

The tissue distribution and the level of the target gene are evaluated for every sample in normal and cancer tissues. Total RNA is extracted from normal tissues, cancer tissues, and from cancers and the corresponding matched adjacent tissues. Subsequently, first strand cDNA is prepared with reverse transcriptase and the polymerase chain reaction is done using primers and Taqman probes specific to each target gene. The results are analyzed using the ABI PRISM 7700 Sequence Detector. The absolute numbers are relative levels of expression of the target gene in a particular tissue compared to the calibrator tissue.

One of ordinary skill can design appropriate primers. The relative levels of expression of the BSNA versus normal tissues and other cancer tissues can then be determined. All the values are compared to a normal tissue (calibrator). These RNA samples are commercially available pools, originated by pooling samples of a particular tissue from different individuals.

The relative levels of expression of the BSNA in pairs of matching samples and 1 cancer and 1 normal/normal adjacent of tissue may also be determined. All the values are compared to a normal tissue (calibrator). A matching pair is formed by mRNA from the cancer sample for a particular tissue and mRNA from the normal adjacent sample for that same tissue from the same individual.

In the analysis of matching samples, BSNAs show a high degree of tissue specificity for the tissue of interest. These results confirm the tissue specificity results obtained with normal pooled samples.

Further, the level of mRNA expression in cancer samples and the isogenic normal adjacent tissue from the same individual are compared. This comparison provides an indication of specificity for the cancer stage (*e.g.* higher levels of mRNA expression in the cancer sample compared to the normal adjacent).

Altogether, the high level of tissue specificity, plus the mRNA overexpression in matching samples tested are indicative of SEQ ID NO: 1 through 171 being diagnostic markers for cancer.

Example 2B: Custom Microarray Experiment

Custom oligonucleotide microarrays were provided by Agilent Technologies, Inc. (Palo Alto, CA). The microarrays were fabricated by Agilent using their technology for the *in-situ* synthesis of 60mer oligonucleotides (Hughes, et al. 2001, Nature Biotechnology 19:342-347). The 60mer microarray probes were designed by Agilent, from gene sequences provided by diaDexus, using Agilent proprietary algorithms. Whenever possible two differents 60mers were designed for each gene of interest.

All microarray experiments were two-color experiments and were performed using Agilent-recommended protocols and reagents. Briefly, each microarray was hybridized with cRNAs synthesized from polyA+ RNA, isolated from cancer and normal tissues, labeled with fluorescent dyes Cyanine3 and Cyanine5 (NEN Life Science Products, Inc., Boston, MA) using a linear amplification method (Agilent). In each experiment, the experimental sample was polyA+ RNA isolated from cancer tissue from

a single individual and the reference sample was a pool of polyA+ RNA isolated from normal tissues of the same organ as the cancerous tissue (*i.e.* normal breast tissue in experiments with breast cancer samples). Hybridizations were carried out at 60°C, overnight using Agilent *in-situ* hybridization buffer. Following washing, arrays were scanned with a GenePix 4000B Microarray Scanner (Axon Instruments, Inc., Union City, CA). The resulting images were analyzed with GenePix Pro 3.0 Microarray Acquisition and Analysis Software (Axon). A total of 36 experiments comparing the expression patterns of breast cancer derived polyA+ RNA (9 stage 1 cancers, 23 stage 2 cancers, 4 stage 3 cancers) to polyA+ RNA isolated from a pool of 10 normal breast tissues were analyzed.

Data normalization and expression profiling were done with Expressionist software from GeneData Inc. (Daly City, CA/Basel, Switzerland). Gene expression analysis was performed using only experiments that meet certain quality criteria. The quality criteria that experiments must meet are a combination of evaluations performed by the Expressionist software and evaluations performed manually using raw and normalized data. To evaluate raw data quality, detection limits (the mean signal for a replicated negative control $\pm$ 2 Standard Deviations (SD)) for each channel were calculated. The detection limit is a measure of non-specific hybridization. Arrays with poor detection limits were not analyzed and the experiments were repeated. To evaluate normalized data quality, positive control elements included in the array were utilized. These array features should have a mean ratio of 1 (no differential expression). If these features have a mean ratio of greater than 1.5-fold up or down, the experiments were not analyzed further and were repeated. In addition to traditional scatter plots demonstrating the distribution of signal in each experiment, the Expressionist software also has minimum thresholding criteria that employs user defined parameters to identify quality data. Only those features that meet the threshhold criteria were included in the filtering and analyses carried out by Expressionist. The thresholding settings employed require a minimum area percentage of 60% [(% pixels > background $\pm$ 2SD)-(% pixels saturated)], and a minimum signal to noise ratio of 2.0 in both channels. By these criteria, very low expressors and saturated features were not included in analysis.

Relative expression data was collected from Expressionist based on meeting the quality parameters described above. Sensitivity data was calculated using an analysis tool. Up- and down- regulated genes were identified using criteria for percentage of valid values obtained, and the percentage of experiments in which the gene is up- or

down-regulated. These criteria were set independently for each data set, depending on the size and the nature of the data set. Results for several BSNAs are shown in the following table. The first three columns of the table contain information about the sequence itself (Oligo ID, Parent ID, and SEQ ID NO), the next 3 columns show the results obtained. '%valid' indicates the percentage of 36 unique experiments total in which a valid expression value was obtained, '%up' indicates the percentage of 20 experiments in which up-regulation of at least 2.5-fold was observed, and '%down' indicates the percentage of the 36 experiments in which down-regulation of at least 2.5-fold was observed. The last column in Table 1 describes the location of the microarray probe (oligo) relative to the sequence.

OligoID	Parent ID	Patent # SEQ ID NO	Sensitivity of up and down regulation			Oligo Seq location in original seq.	Oligo Seq location in FLEX seq
			% valid	% up	% down		
16052	8056	DEX0157_74, DEX0131_52 SEQ ID NO: 127/128	100	11.1	33.3	75-134	1928-1987
24688	5998	DEX0167_22, DEX0157_95, DEX0133_22, DEX0131_78 SEQ ID NO: 160/161	94.4	2.8	58.3	437-496	1093-1152
24689	5998	DEX0157_95, DEX0131_78 SEQ ID NO: 160/161	97.2	2.8	61.1	397-456	
27873	8713	DEX0157_74, DEX0131_52 SEQ ID NO: 127/128	100	13.9	30.6	101-160	1954-2013
33090	5973	DEX0157_73, DEX0131_56 SEQ ID NO: 125/126	97.2	2.8	44.4	408-466	2142-2200
33091	5973	DEX0157_73, DEX0131_56 SEQ ID NO: 125/126	100	2.8	41.7	368-427	1221-1280

Example 3: Protein Expression

The BSNA is amplified by polymerase chain reaction (PCR) and the amplified DNA fragment encoding the BSNA is subcloned in pET-21d for expression in *E. coli*. In addition to the BSNA coding sequence, codons for two amino acids, Met-Ala, flanking the NH_2-terminus of the coding sequence of BSNA, and six histidines, flanking the

COOH-terminus of the coding sequence of BSNA, are incorporated to serve as initiating Met/restriction site and purification tag, respectively.

An over-expressed protein band of the appropriate molecular weight may be observed on a Coomassie blue stained polyacrylamide gel. This protein band is confirmed by Western blot analysis using monoclonal antibody against 6X Histidine tag.

Large-scale purification of BSP was achieved using cell paste generated from 6-liter bacterial cultures, and purified using immobilized metal affinity chromatography (IMAC). Soluble fractions that had been separated from total cell lysate were incubated with a nickle chelating resin. The column was packed and washed with five column volumes of wash buffer. BSP was eluted stepwise with various concentration imidazole buffers.

Example 4: Protein Fusions

Briefly, the human Fc portion of the IgG molecule can be PCR amplified, using primers that span the 5'and 3' ends of the sequence described below. These primers also should have convenient restriction enzyme sites that will facilitate cloning into an expression vector, preferably a mammalian expression vector. For example, if pC4 (Accession No. 209646) is used, the human Fc portion can be ligated into the BamHI cloning site. Note that the 3' BamHI site should be destroyed. Next, the vector containing the human Fc portion is re-restricted with BamHI, linearizing the vector, and a polynucleotide of the present invention, isolated by the PCR protocol described in Example 2, is ligated into this BamHI site. Note that the polynucleotide is cloned without a stop codon, otherwise a fusion protein will not be produced. If the naturally occurring signal sequence is used to produce the secreted protein, pC4 does not need a second signal peptide. Alternatively, if the naturally occurring signal sequence is not used, the vector can be modified to include a heterologous signal sequence. *See, e. g.*, WO 96/34891.

Example 5: Production of an Antibody from a Polypeptide

In general, such procedures involve immunizing an animal (preferably a mouse) with polypeptide or, more preferably, with a secreted polypeptide-expressing cell. Such cells may be cultured in any suitable tissue culture medium; however, it is preferable to culture cells in Earle's modified Eagle's medium supplemented with 10% fetal bovine serum (inactivated at about 56°C), and supplemented with about 10 g/l of nonessential amino acids, about 1,000 U/ml of penicillin, and about 100, µg/ml of streptomycin. The

splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferable to employ the parent myeloma cell line (SP20), available from the ATCC. After fusion, the resulting hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.*, *Gastroenterology* 80: 225-232 (1981).

The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the polypeptide. Alternatively, additional antibodies capable of binding to the polypeptide can be produced in a two-step procedure using anti-idiotypic antibodies. Such a method makes use of the fact that antibodies are themselves antigens, and therefore, it is possible to obtain an antibody which binds to a second antibody. In accordance with this method, protein specific antibodies are used to immunize an animal, preferably a mouse. The splenocytes of such an animal are then used to produce hybridoma cells, and the hybridoma cells are screened to identify clones which produce an antibody whose ability to bind to the protein-specific antibody can be blocked by the polypeptide. Such antibodies comprise anti-idiotypic antibodies to the protein specific antibody and can be used to immunize an animal to induce formation of further protein-specific antibodies. Using the Jameson-Wolf methods the following epitopes were predicted. (Jameson and Wolf, CABIOS, 4(1), 181-186, 1988, the contents of which are incorporated by reference).

The predicted antigenicity for the amino acid sequences is as follows:

DEX ID	ANTIGENICITY Position, AI Ave, Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
DEX0306_172			Myristyl 28-33; 53-58; 60-65;Pkc_Phospho_Site 67-69;	26, .882, .574
DEX0306_173			Myristyl 13-18; Pkc_Phospho_Site 19-21;	
DEX0306_174		1,i20-42o		
DEX0306_175	11-21,1.07,11		Pkc_Phospho_Site 4-6;12-14;	
DEX0306_176	52-69,1.16,18 9-18,1.16,10		Asn_Glycosylation 82-85; Ck2_Phospho_Site 7-10; Myristyl 79-84;	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
			Pkc_Phospho_Site 4-6;	
DEX0306_177			Asn_Glycosylation 7-10;55-58; Ck2_Phospho_Site 22-25;57-60; Pkc_Phospho_Site 57-59; Tyr_Phospho_Site 46-52;	
DEX0306_178	10-47,1.07,38 80-141, 1.03, 62		Myristyl 33-38;129-134; Pkc_Phospho_Site 116-118;147-149;	
DEX0306_179			Myristyl 3-8;	
DEX0306_180	59-74,1.04,16		Ck2_Phospho_Site 4-7;49-52; Myristyl 45-50;50-55;80-85;86-91;95-100; Pkc_Phospho_Site 60-62;65-67;69-71;	
DEX0306_182			Myristyl 22-27;	
DEX0306_184	12-36,1.22,25		Asn_Glycosylation 32-35; Camp_Phospho_Site 26-29; Ck2_Phospho_Site 9-12; Pkc_Phospho_Site 25-27;	
DEX0306_185	6-39,1.13,34		Asn_Glycosylation 64-67; Ck2_Phospho_Site 37-40;65-68; Glycosaminoglycan 48-51; Myristyl 14-19;49-54;51-56; Pkc_Phospho_Site 18-20;42-44;	
DEX0306_187		1,o25-47i	Ck2_Phospho_Site 70-73; Myristyl 7-12; Pkc_Phospho_Site 42-44;	
DEX0306_188		3,i5-22o32-54i61-83o	Myristyl 27-32;141-146;144-149; Pkc_Phospho_Site 17-19;55-57;90-92;111-113;	17, .989, .91
DEX0306_			Ck2_Phospho_Site	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
190			73-76; Myristyl 12-17;17-22;66-71; Pkc_Phospho_Site 91-93;	
DEX0306_192			Pkc_Phospho_Site 6-8;	
DEX0306_193			Myristyl 4-9;	
DEX0306_194	415-439, 1.14,25 242-251, 1.13,10 459-528, 1.11,70 159-197, 1.09,39 777-810, 1.09,34 632-669, 1.07,38 1034-1044, 1.04,11 1077-1103, 1.03,27		Asn_Glycosylation 12-15;19-22;23-26;151-154;513-516;873-876;886-889; Camp_Phospho_Site 107-110; Ck2_Phospho_Site 72-75;260-263;283-286;319-322;463-466;807-810;975-978; Glycosaminoglycan 125-128;905-908;913-916; Myristyl 13-18;28-33;30-35;52-57;53-58;58-63;61-66;62-67;126-131;179-184;372-377;529-534;699-704;716-721;717-722;721-726;837-842;845-850;889-894;906-911;910-915; Pkc_Phospho_Site 129-131;160-162;188-190;189-191;356-358;613-615;822-824;825-827; Prokar_Lipoprotein 44-54;	
DEX0306_195			Pkc_Phospho_Site 6-8;	
DEX0306_196			Pkc_Phospho_Site 24-26;33-35;	
DEX0306_197			Pkc_Phospho_Site 7-9;	
DEX0306_198	39-55,1.09,17 25-34,1.05,10		Ck2_Phospho_Site 92-95; Pkc_Phospho_Site 107-109;	
DEX0306_199	97-113, 1.09,17		Ck2_Phospho_Site 150-153;193-	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
	83-92,1.05,10		196;200-203; Myristyl 11-16;178-183; Pkc_Phospho_Site 165-167; Tyr_Phospho_Site 53-61;	
DEX0306_200		1,i12-34o	Asn_Glycosylation 20-23; Myristyl 18-23;	
DEX0306_201			Myristyl 16-21; Pkc_Phospho_Site 24-26;	24,.944,.779
DEX0306_202	25-37,1.17,13		Ck2_Phospho_Site 12-15; Myristyl 27-32;31-36;53-58;	
DEX0306_203			Asn_Glycosylation 28-31; Myristyl 8-13;62-67;63-68;64-69;	
DEX0306_204			Pkc_Phospho_Site 2-4;	
DEX0306_205			Ck2_Phospho_Site 60-63;77-80; Myristyl 14-19; Pkc_Phospho_Site 57-59;	
DEX0306_206		1,o5-24i	Myristyl 4-9;	
DEX0306_207			Ck2_Phospho_Site 64-67;75-78; Myristyl 71-76;81-86;85-90;	
DEX0306_208			Asn_Glycosylation 53-56;62-65; Myristyl 72-77; Pkc_Phospho_Site 63-65;64-66;	
DEX0306_209			Asn_Glycosylation 47-50; Pkc_Phospho_Site 28-30;38-40; Tyr_Phospho_Site 29-36;30-36;	
DEX0306_211			Asn_Glycosylation 33-36; Ck2_Phospho_Site 17-20; Pkc_Phospho_Site 26-28;	
DEX0306_212	30-39,1.06,10		Ck2_Phospho_Site 76-79; Myristyl 19-24;31-36;92-97; Pkc_Phospho_Site	17,.97,.829

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
			12-14;76-78;	
DEX0306_213			Pkc_Phospho_Site 29-31;	
DEX0306_214			Myristyl 43-48;48-53;	
DEX0306_215	104-118, 1.16,15		Myristyl 90-95;101-106;104-109;	21,.973,.82
DEX0306_216			Ck2_Phospho_Site 5-8;	
DEX0306_217		1,i11-33o	Myristyl 42-47;54-59;67-72; Pkc_Phospho_Site 4-6;37-39;	33,.982,.823
DEX0306_218			Asn_Glycosylation 12-15; Ck2_Phospho_Site 8-11; Myristyl 3-8; Pkc_Phospho_Site 23-25;	
DEX0306_219			Asn_Glycosylation 21-24; Ck2_Phospho_Site 43-46; Pkc_Phospho_Site 23-25;	
DEX0306_220	14-32,1.13,19		Amidation 19-22; Pkc_Phospho_Site 23-25;	
DEX0306_221			Pkc_Phospho_Site 18-20;	
DEX0306_223			Pkc_Phospho_Site 2-4;	
DEX0306_224			Ck2_Phospho_Site 31-34;38-41;57-60;79-82;85-88; Pkc_Phospho_Site 7-9;	
DEX0306_225		1,i7-26o	Asn_Glycosylation 34-37; Ck2_Phospho_Site 36-39;	
DEX0306_226			Pkc_Phospho_Site 34-36;	15,.918,.744
DEX0306_227	52-72,1.19,21	1,i73-95o	Amidation 66-69; Ck2_Phospho_Site 6-9; Myristyl 74-79;78-83;	
DEX0306_228		1,i20-42o		
DEX0306_230		1,o22-44i	Prokar_Lipoprotein 23-33;	
DEX0306_231			Camp_Phospho_Site 3-6; Myristyl 31-	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
			36;90-95;	
DEX0306_232		1,o15-32i	Myristyl 47-52; Pkc_Phospho_Site 2-4;	
DEX0306_233			Asn_Glycosylation 4-7;	
DEX0306_234	24-39,1.2,16		Myristyl 8-13; Pkc_Phospho_Site 65-67;	
DEX0306_235	560-572, 1.27,13 509-519, 1.23,11 1126-1153, 1.19,28 861-873, 1.18,13 794-804, 1.16,11 964-976, 1.16,13 880-901, 1.16,22 812-828, 1.11,17 588-612, 1.09,25 41-77, 1.07,37 461-489, 1.07,29 735-751, 1.07,17 978-1011, 1.06,34 535-558, 1.04,24 1081-1.04,17 620-644, 1.03,25 654-671, 1.01,18 354-382,1,29		Amidation 281-284;403-406;721-724; Asn_Glycosylation 633-636;655-658; Atp_Gtp_A 507-514; Camp_Phospho_Site 54-57;479-482; Ck2_Phospho_Site 132-135;144-147;181-184;209-212;217-220;244-247;310-313;332-335;345-348;546-549;558-561;560-563;593-596;617-620;622-625;635-638;651-654;656-659;697-700;739-742;740-743;745-748;969-972; Glycosaminoglycan 482-485;719-722; Myristyl 110-115;130-135;142-147;159-164;230-235;254-259;277-282;341-346;400-405;510-515;572-577;582-587;645-650;721-726;823-828;842-847;843-848;846-851;872-877;922-927;940-945;954-959; Pkc_Phospho_Site 72-74;83-85;148-150;155-157;156-158;209-211;627-629;635-637;656-658;660-662;661-663;736-738;739-741;745-747;766-768;802-804;813-815;913-915;965-967;973-975;	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
			Tyr_Phospho_Site 55-62;426-433; Zinc_Finger_C2h2 36-56;176-197;250-270;278-298;337-357;517-537;	
DEX0306_236	11-29,1,19	1,o32-54i		
DEX0306_237			Glycosaminoglycan 80-83; Myristyl 14-19;54-59;58-63; Pkc_Phospho_Site 68-70;80-82;	
DEX0306_238		1,o62-84i	Asn_Glycosylation 30-33; Pkc_Phospho_Site 31-33;	
DEX0306_239	42-63,1.12,22		Asn_Glycosylation 145-148; Ck2_Phospho_Site 4-7;63-66;151-154; Euk_Co2_Anhydrase 126-142; Myristyl 25-30;33-38;125-130; Pkc_Phospho_Site 280-282;	
DEX0306_240	20-34,1.08,15		Asn_Glycosylation 53-56; Camp_Phospho_Site 41-44; Pkc_Phospho_Site 39-41;	
DEX0306_242			Myristyl 49-54; Pkc_Phospho_Site 33-35;	
DEX0306_243			Ck2_Phospho_Site 23-26;24-27; Pkc_Phospho_Site 9-11;23-25;	
DEX0306_244			Asn_Glycosylation 4-7;	
DEX0306_245	45-55,1.15,11		Camp_Phospho_Site 51-54; Ck2_Phospho_Site 60-63; Pkc_Phospho_Site 22-24;	
DEX0306_246			Pkc_Phospho_Site 7-9;35-37;	
DEX0306_247			Myristyl 86-91; Pkc_Phospho_Site 17-19;	22,.929,.652

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
DEX0306_248				18,.993,.914
DEX0306_249			Asn_Glycosylation 2-5; Ck2_Phospho_Site 54-57; Pkc_Phospho_Site 54-56;	28,.911,.74
DEX0306_250	142-180, 1.03,39 9-21,1,13		Asn_Glycosylation 13-16;132-135; Ck2_Phospho_Site 97-100; Pkc_Phospho_Site 17-19;55-57;113-115;134-136;153-155;	
DEX0306_251	113-123, 1.14,11 37-60,1.09,24		Camp_Phospho_Site 50-53; Ck2_Phospho_Site 88-91; Pkc_Phospho_Site 39-41;49-51;88-90; Prokar_Lipoprotein 59-69; Tyr_Phospho_Site 87-95;	
DEX0306_252			Pkc_Phospho_Site 10-12;	
DEX0306_253		1,i12-43o	Myristyl 30-35; Prokar_Lipoprotein 12-22;	30,.996,.862
DEX0306_254			Ck2_Phospho_Site 16-19; Myristyl 31-36;36-41; Pkc_Phospho_Site 32-34; Rgd 25-27;	
DEX0306_255			Asn_Glycosylation 386-389;516-519;536-539;626-629;638-641;883-886; Camp_Phospho_Site 61-64; Ck2_Phospho_Site 147-150;201-204;205-208;252-255;394-397;435-438;462-465;491-494;511-514;524-527;552-555;632-635;646-649;756-759;839-842;867-870;887-890; Myristyl 25-30;263-268;751-	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
			756;879-884; Pkc_Phospho_Site 29-31;107-109;147-149;201-203;506-508; Tyr_Phospho_Site 467-473;	
DEX0306_256	65-75,1.02,11 25-50,1.02,26		Asn_Glycosylation 56-59; Myristyl 14-19; Prokar_Lipoprotein 8-18;	
DEX0306_257	179-203, 1.18,25 527-569, 1.15,43 422-464, 1.11,43 20-39,1.06,20 335-367, 1.06,33 43-117, 1.01,75		Amidation 267-270; Asn_Glycosylation 176-179; Camp_Phospho_Site 71-74;324-327; Ck2_Phospho_Site 42-45;54-57;75-78;99-102;109-112;161-164;197-200;206-209;223-226;228-231;273-276;283-286;336-339;447-450;482-485;497-500;567-570; Glycosaminoglycan 246-249; Myristyl 24-29;38-43;86-91;124-129;249-254;262-267;278-283;290-295;332-337;410-415;430-435; Pkc_Phospho_Site 12-14;18-20;28-30;35-37;54-56;69-71;296-298;336-338;411-413;434-436; Tyr_Phospho_Site 23-29;137-144;310-318;	
DEX0306_258			Ck2_Phospho_Site 34-37;	
DEX0306_259			Asn_Glycosylation 31-34;	
DEX0306_260			Camp_Phospho_Site 6-9; Myristyl 54-59;	
DEX0306_261	96-105, 1.19,10		Ck2_Phospho_Site 71-74;101-104; Glycosaminoglycan 55-58; Myristyl 52-57;54-59;58-	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
			63;67-72; Pkc_Phospho_Site 17-19;137-139;146-148;197-199;215-217; Prokar_Lipoprotein 164-174;	
DEX0306_262	30-41,1.02,12		Asn_Glycosylation 86-89; Ck2_Phospho_Site 21-24; Myristyl 96-101; Pkc_Phospho_Site 18-20;	
DEX0306_263	239-249, 1.13,11		Amidation 72-75; Asn_Glycosylation 119-122;120-123; Camp_Phospho_Site 107-110;216-219; Ck2_Phospho_Site 28-31;43-46;63-66;160-163;169-172;187-190; Myristyl 69-74;158-163; Pkc_Phospho_Site 17-19;24-26;35-37;52-54;59-61;106-108;122-124;184-186; Prokar_ Lipoprotein 248-258;	
DEX0306_264			Myristyl 35-40; Pkc_Phospho_Site 21-23;22-24;	
DEX0306_265		1,i7-29o	Camp_Phospho_Site 47-50; Ck2_Phospho_Site 54-57; Myristyl 37-42; Pkc_Phospho_Site 72-74;	
DEX0306_266			Asn_Glycosylation 7-10;17-20; Pkc_Phospho_Site 2-4;	
DEX0306_267			Amidation 43-46; Ck2_Phospho_Site 79-82; Pkc_Phospho_Site 11-13;89-91;	
DEX0306_268			Pkc_Phospho_Site 8-10;45-47; Prokar_Lipoprotein 32-42;	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
DEX0306_269			Camp_Phospho_Site 66-69; Ck2_Phospho_Site 12-15;34-37;56-59; Myristyl 30-35; Pkc_Phospho_Site 34-36;56-58;	
DEX0306_270	49-134,1,86		Asn_Glycosylation 46-49; Ck2_Phospho_Site 65-68;84-87;93-96;109-112; Myristyl 4-9;59-64; Pkc_Phospho_Site 60-62;89-91;104-106;115-117;116-118; Tyr_Phospho_Site 92-99;117-124;118-124;	
DEX0306_272	235-299,1.1,65 369-406,1.07,38 99-109,1.01,11		Asn_Glycosylation 37-40;69-72;284-287; Ck2_Phospho_Site 85-88;141-144;149-152;192-195;204-207; Glycosaminoglycan 433-436; Myristyl 43-48;44-49;96-101;118-123;402-407;406-411;432-437;438-443; Pkc_Phospho_Site 48-50;433-435; Rgd 278-280; Tyr_Phospho_Site 50-56;	
DEX0306_273			Pkc_Phospho_Site 6-8;15-17;	
DEX0306_274			Asn_Glycosylation 44-47;	
DEX0306_275			Asn_Glycosylation 78-81; Ck2_Phospho_Site 17-20; Myristyl 13-18;	
DEX0306_276			Ck2_Phospho_Site 58-61; Glycosaminoglycan 93-96; Myristyl 28-33;48-53;50-55;67-72;71-76; Pkc_Phospho_Site	

DEX ID	ANTIGENICITY Position, AI Ave, Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
			5-7;18-20;44-46;57-59; Rgd 59-61;	
DEX0306_277		1,o37-59i	Ck2_Phospho_Site 22-25; Myristyl 71-76; Pkc_Phospho_Site 12-14;	
DEX0306_279			Myristyl 15-20;	
DEX0306_280			Ck2_Phospho_Site 76-79; Pkc_Phospho_Site 16-18;	27,.985,.682
DEX0306_281	17-29,1.07,13		Myristyl 5-10;9-14; Pkc_Phospho_Site 24-26;	
DEX0306_282		1,o15-32i		
DEX0306_283			Asn_Glycosylation 35-38; Ck2_Phospho_Site 37-40; Myristyl 3-8; Pkc_Phospho_Site 57-59;	
DEX0306_284	28-37,1.09,10		Ck2_Phospho_Site 46-49; Pkc_Phospho_Site 32-34;	21,.958,.821
DEX0306_285	226-245, 1.37,20 489-501, 1.22,13 1271-1284, 1.21,14 1192-1203, 1.11,12 745-755, 1.09,11 929-940, 1.08,12 1039-1051, 1.08,13 1133-1150, 1.05,18 547-576, 1.05,30 89-98,1.04,10 22-53,1.03,32 1073-1086, 1.03,14 1243-1253, 1.03,11 1418-1461,		Amidation 473-476; Asn_Glycosylation 512-515;726-729; Camp_Phospho_Site 475-478;571-574;646-649; Ck2_Phospho_Site 29-32;143-146;176-179;228-231;230-233;232-235;263-266;294-297;388-391;447-450;493-496;506-509;517-520;581-584;664-667;890-893;929-932; Gram_Pos_Anchoring 670-675; Myristyl 49-54;56-61;125-130;152-157;185-190;214-219;677-682;708-713;840-845;921-926;	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
	1.01,44		Pkc_Phospho_Site 21-23;29-31;143-145;388-390;415-417;443-445;530-532;539-541;552-554;565-567;581-583;748-750;802-804;925-927;931-933;987-989;996-998; Tyr_Phospho_Site 867-874;Amidation 473-476; Asn_Glycosylation 512-515;726-729; Camp_Phospho_Site 475-478;571-574;646-649; Ck2_Phospho_Site 29-32;143-146;176-179;228-231;230-233;232-235;263-266;294-297;388-391;447-450;493-496;506-509;517-520;581-584;664-667;890-893;929-932; Gram_Pos_ Anchoring 670-675; Myristyl 49-54;56-61;125-130;152-157;185-190;214-219;677-682;708-713;840-845;921-926; Pkc_Phospho_Site 21-23;29-31;143-145;388-390;415-417;443-445;530-532;539-541;552-554;565-567;581-583;748-750;802-804;925-927;931-933;987-989;996-998; Tyr_Phospho_Site 867-874;	
DEX0306_286		2,i13-30o35-54i	Asn_Glycosylation 15-18; Ck2_Phospho_Site 41-44; Myristyl 2-7; Pkc_Phospho_Site 6-8;	
DEX0306_			Asn_Glycosylation	

DEX ID	ANTIGENICITY Position, AI Ave,Length	TRANSMEMBRANE Predicted Helix, Topology	PTM PTM	SIGNAL PEPTIDE Position, Max Score, Mean Score
287			43-46;51-54; Ck2_Phospho_Site 34-37; Pkc_Phospho_Site 70-72;	
DEX0306_288			Asn_Glycosylation 42-45; Camp_Phospho_Site 12-15; Myristyl 4-9;	
DEX0306_290	20-31,1.14,12		Pkc_Phospho_Site 6-8;21-23;	
DEX0306_291			Glycosaminoglycan 31-34; Myristyl 30-35;	
DEX0306_292			Camp_Phospho_Site 8-11; Ck2_Phospho_Site 11-14;	
DEX0306_293			Ck2_Phospho_Site 36-39; Myristyl 2-7;94-99;	
DEX0306_294	31-52,1.01,22		Pkc_Phospho_Site 47-49;	
DEX0306_295			Myristyl 56-61;	

Example 6: Method of Determining Alterations in a Gene Corresponding to a Polynucleotide

RNA is isolated from individual patients or from a family of individuals that have a phenotype of interest. cDNA is then generated from these RNA samples using protocols known in the art. *See*, Sambrook (2001), *supra*. The cDNA is then used as a template for PCR, employing primers surrounding regions of interest in SEQ ID NO: 1 through 171. Suggested PCR conditions consist of 35 cycles at 95°C for 30 seconds; 60-120 seconds at 52-58°C; and 60-120 seconds at 70°C, using buffer solutions described in Sidransky *et al., Science* 252(5006): 706-9 (1991). *See also* Sidransky *et al., Science* 278(5340): 1054-9 (1997).

PCR products are then sequenced using primers labeled at their 5' end with T4 polynucleotide kinase, employing SequiTherm Polymerase. (Epicentre Technologies). The intron-exon borders of selected exons is also determined and genomic PCR products analyzed to confirm the results. PCR products harboring suspected mutations are then cloned and sequenced to validate the results of the direct sequencing. PCR products is

cloned into T-tailed vectors as described in Holton *et al., Nucleic Acids Res.*, 19: 1156 (1991) and sequenced with T7 polymerase (United States Biochemical). Affected individuals are identified by mutations not present in unaffected individuals.

Genomic rearrangements may also be determined. Genomic clones are nick-translated with digoxigenin deoxyuridine 5' triphosphate (Boehringer Manheim), and FISH is performed as described in Johnson *et al., Methods Cell Biol.* 35: 73-99 (1991). Hybridization with the labeled probe is carried out using a vast excess of human cot-1 DNA for specific hybridization to the corresponding genomic locus.

Chromosomes are counterstained with 4,6-diamino-2-phenylidole and propidium iodide, producing a combination of C-and R-bands. Aligned images for precise mapping are obtained using a triple-band filter set (Chroma Technology, Brattleboro, VT) in combination with a cooled charge-coupled device camera (Photometrics, Tucson, AZ) and variable excitation wavelength filters. *Id.* Image collection, analysis and chromosomal fractional length measurements are performed using the ISee Graphical Program System. (Inovision Corporation, Durham, NC.) Chromosome alterations of the genomic region hybridized by the probe are identified as insertions, deletions, and translocations. These alterations are used as a diagnostic marker for an associated disease.

Example 7: Method of Detecting Abnormal Levels of a Polypeptide in a Biological Sample

Antibody-sandwich ELISAs are used to detect polypeptides in a sample, preferably a biological sample. Wells of a microtiter plate are coated with specific antibodies, at a final concentration of 0.2 to 10 μg/ml. The antibodies are either monoclonal or polyclonal and are produced by the method described above. The wells are blocked so that non-specific binding of the polypeptide to the well is reduced. The coated wells are then incubated for > 2 hours at RT with a sample containing the polypeptide. Preferably, serial dilutions of the sample should be used to validate results. The plates are then washed three times with deionized or distilled water to remove unbound polypeptide. Next, 50 μl of specific antibody-alkaline phosphatase conjugate, at a concentration of 25-400 ng, is added and incubated for 2 hours at room temperature. The plates are again washed three times with deionized or distilled water to remove unbound conjugate. 75 μl of 4-methylumbelliferyl phosphate (MUP) or p-nitrophenyl

phosphate (NPP) substrate solution are added to each well and incubated 1 hour at room temperature.

The reaction is measured by a microtiter plate reader. A standard curve is prepared, using serial dilutions of a control sample, and polypeptide concentrations are plotted on the X-axis (log scale) and fluorescence or absorbance on the Y-axis (linear scale). The concentration of the polypeptide in the sample is calculated using the standard curve.

Example 8: Formulating a Polypeptide

The secreted polypeptide composition will be formulated and dosed in a fashion consistent with good medical practice, taking into account the clinical condition of the individual patient (especially the side effects of treatment with the secreted polypeptide alone), the site of delivery, the method of administration, the scheduling of administration, and other factors known to practitioners. The "effective amount" for purposes herein is thus determined by such considerations.

As a general proposition, the total pharmaceutically effective amount of secreted polypeptide administered parenterally per dose will be in the range of about 1 , μg/kg/day to 10 mg/kg/day of patient body weight, although, as noted above, this will be subject to therapeutic discretion. More preferably, this dose is at least 0.01 mg/kg/day, and most preferably for humans between about 0.01 and 1 mg/kg/day for the hormone. If given continuously, the secreted polypeptide is typically administered at a dose rate of about 1 μg/kg/hour to about 50 mg/kg/hour, either by 1-4 injections per day or by continuous subcutaneous infusions, for example, using a mini-pump. An intravenous bag solution may also be employed. The length of treatment needed to observe changes and the interval following treatment for responses to occur appears to vary depending on the desired effect.

Pharmaceutical compositions containing the secreted protein of the invention are administered orally, rectally, parenterally, intracistemally, intravaginally, intraperitoneally, topically (as by powders, ointments, gels, drops or transdermal patch), bucally, or as an oral or nasal spray. "Pharmaceutically acceptable carrier" refers to a non-toxic solid, semisolid or liquid filler, diluent, encapsulating material or formulation auxiliary of any type. The term "parenteral" as used herein refers to modes of administration which include intravenous, intramuscular, intraperitoneal, intrasternal, subcutaneous and intraarticular injection and infusion.

The secreted polypeptide is also suitably administered by sustained-release systems. Suitable examples of sustained-release compositions include semipermeable polymer matrices in the form of shaped articles, e. g., films, or microcapsules. Sustained-release matrices include polylactides (U. S. Pat. No.3,773,919, EP 58,481), copolymers of L-glutamic acid and gamma-ethyl-L-glutamate (Sidman, U. et al., Biopolymers 22: 547-556 (1983)), poly (2-hydroxyethyl methacrylate) (R. Langer et al., J. Biomed. Mater. Res. 15: 167-277 (1981), and R. Langer, Chem. Tech. 12: 98-105 (1982)), ethylene vinyl acetate (R. Langer et al.) or poly-D- (-)-3-hydroxybutyric acid (EP 133,988). Sustained-release compositions also include liposomally entrapped polypeptides. Liposomes containing the secreted polypeptide are prepared by methods known per se: DE Epstein et al., Proc. Natl. Acad. Sci. USA 82: 3688-3692 (1985); Hwang et al., Proc. Natl. Acad. Sci. USA 77: 4030-4034 (1980); EP 52,322; EP 36,676; EP 88,046; EP 143,949; EP 142,641; Japanese Pat. Appl. 83-118008; U. S. Pat. Nos. 4,485,045 and 4,544,545; and EP 102,324. Ordinarily, the liposomes are of the small (about 200-800 Angstroms) unilamellar type in which the lipid content is greater than about 30 mol. percent cholesterol, the selected proportion being adjusted for the optimal secreted polypeptide therapy.

For parenteral administration, in one embodiment, the secreted polypeptide is formulated generally by mixing it at the desired degree of purity, in a unit dosage injectable form (solution, suspension, or emulsion), with a pharmaceutically acceptable carrier, I. e., one that is non-toxic to recipients at the dosages and concentrations employed and is compatible with other ingredients of the formulation.

For example, the formulation preferably does not include oxidizing agents and other compounds that are known to be deleterious to polypeptides. Generally, the formulations are prepared by contacting the polypeptide uniformly and intimately with liquid carriers or finely divided solid carriers or both. Then, if necessary, the product is shaped into the desired formulation. Preferably the carrier is a parenteral carrier, more preferably a solution that is isotonic with the blood of the recipient. Examples of such carrier vehicles include water, saline, Ringer's solution, and dextrose solution. Non-aqueous vehicles such as fixed oils and ethyl oleate are also useful herein, as well as liposomes.

The carrier suitably contains minor amounts of additives such as substances that enhance isotonicity and chemical stability. Such materials are non-toxic to recipients at the dosages and concentrations employed, and include buffers such as phosphate, citrate,

succinate, acetic acid, and other organic acids or their salts; antioxidants such as ascorbic acid; low molecular weight (less than about ten residues) polypeptides, e. g., polyarginine or tripeptides; proteins, such as serum albumin, gelatin, or immunoglobulins; hydrophilic polymers such as polyvinylpyrrolidone; amino acids, such as glycine, glutamic acid, aspartic acid, or arginine; monosaccharides, disaccharides, and other carbohydrates including cellulose or its derivatives, glucose, manose, or dextrins; chelating agents such as EDTA; sugar alcohols such as mannitol or sorbitol; counterions such as sodium; and/or nonionic surfactants such as polysorbates, poloxamers, or PEG.

The secreted polypeptide is typically formulated in such vehicles at a concentration of about 0.1 mg/ml to 100 mg/ml, preferably 1-10 mg/ml, at a pH of about 3 to 8. It will be understood that the use of certain of the foregoing excipients, carriers, or stabilizers will result in the formation of polypeptide salts.

Any polypeptide to be used for therapeutic administration can be sterile. Sterility is readily accomplished by filtration through sterile filtration membranes (e. g., 0.2 micron membranes). Therapeutic polypeptide compositions generally are placed into a container having a sterile access port, for example, an intravenous solution bag or vial having a stopper pierceable by a hypodermic injection needle.

Polypeptides ordinarily will be stored in unit or multi-dose containers, for example, sealed ampules or vials, as an aqueous solution or as a lyophilized formulation for reconstitution. As an example of a lyophilized formulation, 10-ml vials are filled with 5 ml of sterile-filtered 1 % (w/v) aqueous polypeptide solution, and the resulting mixture is lyophilized. The infusion solution is prepared by reconstituting the lyophilized polypeptide using bacteriostatic Water-for-Injection.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Associated with such container (s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval by the agency of manufacture, use or sale for human administration. In addition, the polypeptides of the present invention may be employed in conjunction with other therapeutic compounds.

Example 9: Method of Treating Decreased Levels of the Polypeptide

It will be appreciated that conditions caused by a decrease in the standard or normal expression level of a secreted protein in an individual can be treated by administering the polypeptide of the present invention, preferably in the secreted form.

Thus, the invention also provides a method of treatment of an individual in need of an increased level of the polypeptide comprising administering to such an individual a pharmaceutical composition comprising an amount of the polypeptide to increase the activity level of the polypeptide in such an individual.

For example, a patient with decreased levels of a polypeptide receives a daily dose 0.1-100 μg/kg of the polypeptide for six consecutive days. Preferably, the polypeptide is in the secreted form. The exact details of the dosing scheme, based on administration and formulation, are provided above.

Example 10: Method of Treating Increased Levels of the Polypeptide

Antisense technology is used to inhibit production of a polypeptide of the present invention. This technology is one example of a method of decreasing levels of a polypeptide, preferably a secreted form, due to a variety of etiologies, such as cancer.

For example, a patient diagnosed with abnormally increased levels of a polypeptide is administered intravenously antisense polynucleotides at 0.5, 1.0, 1.5, 2.0 and 3.0 mg/kg day for 21 days. This treatment is repeated after a 7-day rest period if the treatment was well tolerated. The formulation of the antisense polynucleotide is provided above.

Example 11: Method of Treatment Using Gene Therapy

One method of gene therapy transplants fibroblasts, which are capable of expressing a polypeptide, onto a patient. Generally, fibroblasts are obtained from a subject by skin biopsy. The resulting tissue is placed in tissue-culture medium and separated into small pieces. Small chunks of the tissue are placed on a wet surface of a tissue culture flask, approximately ten pieces are placed in each flask. The flask is turned upside down, closed tight and left at room temperature over night. After 24 hours at room temperature, the flask is inverted and the chunks of tissue remain fixed to the bottom of the flask and fresh media (e. g., Ham's F12 media, with 10% FBS, penicillin and streptomycin) is added. The flasks are then incubated at 37°C for approximately one week.

At this time, fresh media is added and subsequently changed every several days. After an additional two weeks in culture, a monolayer of fibroblasts emerge. The monolayer is trypsinized and scaled into larger flasks. pMV-7 (Kirschmeier, P. T. et al., DNA, 7: 219-25 (1988)), flanked by the long terminal repeats of the Moloney murine sarcoma virus, is digested with EcoRI and HindIII and subsequently treated with calf

intestinal phosphatase. The linear vector is fractionated on agarose gel and purified, using glass beads.

The cDNA encoding a polypeptide of the present invention can be amplified using PCR primers which correspond to the 5'and 3'end sequences respectively as set forth in Example 1. Preferably, the 5'primer contains an EcoRI site and the 3'primer includes a HindIII site. Equal quantities of the Moloney murine sarcoma virus linear backbone and the amplified EcoRI and HindIII fragment are added together, in the presence of T4 DNA ligase. The resulting mixture is maintained under conditions appropriate for ligation of the two fragments. The ligation mixture is then used to transform bacteria HB 101, which are then plated onto agar containing kanamycin for the purpose of confirming that the vector has the gene of interest properly inserted.

The amphotropic pA317 or GP+aml2 packaging cells are grown in tissue culture to confluent density in Dulbecco's Modified Eagles Medium (DMEM) with 10% calf serum (CS), penicillin and streptomycin. The MSV vector containing the gene is then added to the media and the packaging cells transduced with the vector. The packaging cells now produce infectious viral particles containing the gene (the packaging cells are now referred to as producer cells).

Fresh media is added to the transduced producer cells, and subsequently, the media is harvested from a 10 cm plate of confluent producer cells. The spent media, containing the infectious viral particles, is filtered through a millipore filter to remove detached producer cells and this media is then used to infect fibroblast cells. Media is removed from a sub-confluent plate of fibroblasts and quickly replaced with the media from the producer cells. This media is removed and replaced with fresh media.

If the titer of virus is high, then virtually all fibroblasts will be infected and no selection is required. If the titer is very low, then it is necessary to use a retroviral vector that has a selectable marker, such as neo or his. Once the fibroblasts have been efficiently infected, the fibroblasts are analyzed to determine whether protein is produced.

The engineered fibroblasts are then transplanted onto the host, either alone or after having been grown to confluence on cytodex 3 microcarrier beads.

Example 12: Method of Treatment Using Gene Therapy-*In Vivo*

Another aspect of the present invention is using *in vivo* gene therapy methods to treat disorders, diseases and conditions. The gene therapy method relates to the introduction of naked nucleic acid (DNA, RNA, and antisense DNA or RNA) sequences into an animal to increase or decrease the expression of the polypeptide.

The polynucleotide of the present invention may be operatively linked to a promoter or any other genetic elements necessary for the expression of the polypeptide by the target tissue. Such gene therapy and delivery techniques and methods are known in the art, see, for example, W0 90/11092, W0 98/11779; U. S. Patent 5,693,622; 5,705,151; 5,580,859; Tabata H. et al. (1997) Cardiovasc. Res. 35 (3): 470-479, Chao J et al. (1997) Pharmacol. Res. 35 (6): 517-522, Wolff J. A. (1997) Neuromuscul. Disord. 7 (5): 314-318, Schwartz B. et al. (1996) Gene Ther. 3 (5): 405-411, Tsurumi Y. et al. (1996) Circulation 94 (12): 3281-3290 (incorporated herein by reference).

The polynucleotide constructs may be delivered by any method that delivers injectable materials to the cells of an animal, such as, injection into the interstitial space of tissues (heart, muscle, skin, lung, liver, intestine and the like). The polynucleotide constructs can be delivered in a pharmaceutically acceptable liquid or aqueous carrier.

The term "naked" polynucleotide, DNA or RNA, refers to sequences that are free from any delivery vehicle that acts to assist, promote, or facilitate entry into the cell, including viral sequences, viral particles, liposome formulations, lipofectin or precipitating agents and the like. However, the polynucleotides of the present invention may also be delivered in liposome formulations (such as those taught in Felgner P. L. et al. (1995) Ann. NY Acad. Sci. 772: 126-139 and Abdallah B. et al. (1995) Biol. Cell 85 (1): 1-7) which can be prepared by methods well known to those skilled in the art.

The polynucleotide vector constructs used in the gene therapy method are preferably constructs that will not integrate into the host genome nor will they contain sequences that allow for replication. Any strong promoter known to those skilled in the art can be used for driving the expression of DNA. Unlike other gene therapies techniques, one major advantage of introducing naked nucleic acid sequences into target cells is the transitory nature of the polynucleotide synthesis in the cells. Studies have shown that non-replicating DNA sequences can be introduced into cells to provide production of the desired polypeptide for periods of up to six months.

The polynucleotide construct can be delivered to the interstitial space of tissues within the an animal, including of muscle, skin, brain, lung, liver, spleen, bone marrow, thymus, heart, lymph, blood, bone, cartilage, pancreas, kidney, gall bladder, stomach, intestine, testis, ovary, uterus, rectum, nervous system, eye, gland, and connective tissue. Interstitial space of the tissues comprises the intercellular fluid, mucopolysaccharide matrix among the reticular fibers of organ tissues, elastic fibers in the walls of vessels or chambers, collagen fibers of fibrous tissues, or that same matrix within connective tissue

ensheathing muscle cells or in the lacunae of bone. It is similarly the space occupied by the plasma of the circulation and the lymph fluid of the lymphatic channels. Delivery to the interstitial space of muscle tissue is preferred for the reasons discussed below. They may be conveniently delivered by injection into the tissues comprising these cells. They are preferably delivered to and expressed in persistent, non-dividing cells which are differentiated, although delivery and expression may be achieved in non-differentiated or less completely differentiated cells, such as, for example, stem cells of blood or skin fibroblasts. *In vivo* muscle cells are particularly competent in their ability to take up and express polynucleotides.

For the naked polynucleotide injection, an effective dosage amount of DNA or RNA will be in the range of from about 0.05 μg/kg body weight to about 50 mg/kg body weight. Preferably the dosage will be from about 0.005 mg/kg to about 20 mg/kg and more preferably from about 0.05 mg/kg to about 5 mg/kg. Of course, as the artisan of ordinary skill will appreciate, this dosage will vary according to the tissue site of injection. The appropriate and effective dosage of nucleic acid sequence can readily be determined by those of ordinary skill in the art and may depend on the condition being treated and the route of administration. The preferred route of administration is by the parenteral route of injection into the interstitial space of tissues. However, other parenteral routes may also be used, such as, inhalation of an aerosol formulation particularly for delivery to lungs or bronchial tissues, throat or mucous membranes of the nose. In addition, naked polynucleotide constructs can be delivered to arteries during angioplasty by the catheter used in the procedure.

The dose response effects of injected polynucleotide in muscle *in vivo* is determined as follows. Suitable template DNA for production of mRNA coding for polypeptide of the present invention is prepared in accordance with a standard recombinant DNA methodology. The template DNA, which may be either circular or linear, is either used as naked DNA or complexed with liposomes. The quadriceps muscles of mice are then injected with various amounts of the template DNA.

Five to six week old female and male Balb/C mice are anesthetized by intraperitoneal injection with 0.3 ml of 2.5% Avertin. A 1.5 cm incision is made on the anterior thigh, and the quadriceps muscle is directly visualized. The template DNA is injected in 0.1 ml of carrier in a 1 cc syringe through a 27 gauge needle over one minute, approximately 0.5 cm from the distal insertion site of the muscle into the knee and about

0.2 cm deep. A suture is placed over the injection site for future localization, and the skin is closed with stainless steel clips.

After an appropriate incubation time (e. g., 7 days) muscle extracts are prepared by excising the entire quadriceps. Every fifth 15 um cross-section of the individual quadriceps muscles is histochemically stained for protein expression. A time course for protein expression may be done in a similar fashion except that quadriceps from different mice are harvested at different times. Persistence of DNA in muscle following injection may be determined by Southern blot analysis after preparing total cellular DNA and HIRT supernatants from injected and control mice.

The results of the above experimentation in mice can be use to extrapolate proper dosages and other treatment parameters in humans and other animals using naked DNA.

Example 13: Transgenic Animals

The polypeptides of the invention can also be expressed in transgenic animals. Animals of any species, including, but not limited to, mice, rats, rabbits, hamsters, guinea pigs, pigs, micro-pigs, goats, sheep, cows and non-human primates, e. g., baboons, monkeys, and chimpanzees may be used to generate transgenic animals. In a specific embodiment, techniques described herein or otherwise known in the art, are used to express polypeptides of the invention in humans, as part of a gene therapy protocol.

Any technique known in the art may be used to introduce the transgene (i. e., polynucleotides of the invention) into animals to produce the founder lines of transgenic animals. Such techniques include, but are not limited to, pronuclear microinjection (Paterson et al., Appl. Microbiol. Biotechnol. 40: 691-698 (1994); Carver et al., Biotechnology (NY) 11: 1263-1270 (1993); Wright et al., Biotechnology (NY) 9: 830-834 (1991); and Hoppe et al., U. S. Patent 4,873,191 (1989)); retrovirus mediated gene transfer into germ lines (Van der Putten et al., Proc. Natl. Acad. Sci., USA 82: 6148-6152 (1985)), blastocysts or embryos; gene targeting in embryonic stem cells (Thompson et al., Cell 56: 313-321 (1989)); electroporation of cells or embryos (Lo, 1983, Mol Cell. Biol. 3: 1803-1814 (1983)); introduction of the polynucleotides of the invention using a gene gun (see, e. g., Ulmer et al., Science 259: 1745 (1993); introducing nucleic acid constructs into embryonic pleuripotent stem cells and transferring the stem cells back into the blastocyst; and sperm mediated gene transfer (Lavitrano et al., Cell 57: 717-723 (1989); etc. For a review of such techniques, see Gordon,"Transgenic Animals," Intl. Rev. Cytol. 115: 171-229 (1989), which is incorporated by reference herein in its entirety.

Any technique known in the art may be used to produce transgenic clones containing polynucleotides of the invention, for example, nuclear transfer into enucleated oocytes of nuclei from cultured embryonic, fetal, or adult cells induced to quiescence (Campell et al., Nature 380: 64-66 (1996); Wilmut et al., Nature 385: 810813 (1997)).

The present invention provides for transgenic animals that carry the transgene in all their cells, as well as animals which carry the transgene in some, but not all their cells, I. e., mosaic animals or chimeric. The transgene may be integrated as a single transgene or as multiple copies such as in concatamers, e. g., head-to-head tandems or head-to-tail tandems. The transgene may also be selectively introduced into and activated in a particular cell type by following, for example, the teaching of Lasko et al. (Lasko et al., Proc. Natl. Acad. Sci. USA 89: 6232-6236 (1992)). The regulatory sequences required for such a cell-type specific activation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art. When it is desired that the polynucleotide transgene be integrated into the chromosomal site of the endogenous gene, gene targeting is preferred. Briefly, when such a technique is to be utilized, vectors containing some nucleotide sequences homologous to the endogenous gene are designed for the purpose of integrating, via homologous recombination with chromosomal sequences, into and disrupting the function of the nucleotide sequence of the endogenous gene. The transgene may also be selectively introduced into a particular cell type, thus inactivating the endogenous gene in only that cell type, by following, for example, the teaching of Gu et al. (Gu et al., Science 265: 103-106 (1994)). The regulatory sequences required for such a cell-type specific inactivation will depend upon the particular cell type of interest, and will be apparent to those of skill in the art.

Once transgenic animals have been generated, the expression of the recombinant gene may be assayed utilizing standard techniques. Initial screening may be accomplished by Southern blot analysis or PCR techniques to analyze animal tissues to verify that integration of the transgene has taken place. The level of mRNA expression of the transgene in the tissues of the transgenic animals may also be assessed using techniques which include, but are not limited to, Northern blot analysis of tissue samples obtained from the animal, in situ hybridization analysis, and reverse transcriptase-PCR (rt-PCR). Samples of transgenic gene-expressing tissue may also be evaluated immunocytochemically or immunohistochemically using antibodies specific for the transgene product.

Once the founder animals are produced, they may be bred, inbred, outbred, or crossbred to produce colonies of the particular animal. Examples of such breeding strategies include, but are not limited to: outbreeding of founder animals with more than one integration site in order to establish separate lines; inbreeding of separate lines in order to produce compound transgenics that express the transgene at higher levels because of the effects of additive expression of each transgene; crossing of heterozygous transgenic animals to produce animals homozygous for a given integration site in order to both augment expression and eliminate the need for screening of animals by DNA analysis; crossing of separate homozygous lines to produce compound heterozygous or homozygous lines; and breeding to place the transgene on a distinct background that is appropriate for an experimental model of interest.

Transgenic animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

Example 14: Knock-Out Animals

Endogenous gene expression can also be reduced by inactivating or "knocking out" the gene and/or its promoter using targeted homologous recombination. (E. g., see Smithies et al., Nature 317: 230-234 (1985); Thomas & Capecchi, Cell 51: 503512 (1987); Thompson et al., Cell 5: 313-321 (1989); each of which is incorporated by reference herein in its entirety). For example, a mutant, non-functional polynucleotide of the invention (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous polynucleotide sequence (either the coding regions or regulatory regions of the gene) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express polypeptides of the invention *in vivo*. In another embodiment, techniques known in the art are used to generate knockouts in cells that contain, but do not express the gene of interest. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the targeted gene. Such approaches are particularly suited in research and agricultural fields where modifications to embryonic stem cells can be used to generate animal offspring with an inactive targeted gene (e. g., see Thomas & Capecchi 1987 and Thompson 1989, supra). However this approach can be routinely adapted for use in humans provided the

recombinant DNA constructs are directly administered or targeted to the required site *in vivo* using appropriate viral vectors that will be apparent to those of skill in the art.

In further embodiments of the invention, cells that are genetically engineered to express the polypeptides of the invention, or alternatively, that are genetically engineered not to express the polypeptides of the invention (e. g., knockouts) are administered to a patient *in vivo*. Such cells may be obtained from the patient (I. e., animal, including human) or an MHC compatible donor and can include, but are not limited to fibroblasts, bone marrow cells, blood cells (e. g., lymphocytes), adipocytes, muscle cells, endothelial cells etc. The cells are genetically engineered *in vitro* using recombinant DNA techniques to introduce the coding sequence of polypeptides of the invention into the cells, or alternatively, to disrupt the coding sequence and/or endogenous regulatory sequence associated with the polypeptides of the invention, e. g., by transduction (using viral vectors, and preferably vectors that integrate the transgene into the cell genome) or transfection procedures, including, but not limited to, the use of plasmids, cosmids, YACs, naked DNA, electroporation, liposomes, etc.

The coding sequence of the polypeptides of the invention can be placed under the control of a strong constitutive or inducible promoter or promoter/enhancer to achieve expression, and preferably secretion, of the polypeptides of the invention. The engineered cells which express and preferably secrete the polypeptides of the invention can be introduced into the patient systemically, e. g., in the circulation, or intraperitoneally.

Alternatively, the cells can be incorporated into a matrix and implanted in the body, e. g., genetically engineered fibroblasts can be implanted as part of a skin graft; genetically engineered endothelial cells can be implanted as part of a lymphatic or vascular graft. (See, for example, Anderson et al. U. S. Patent 5,399,349; and Mulligan & Wilson, U. S. Patent 5,460,959 each of which is incorporated by reference herein in its entirety).

When the cells to be administered are non-autologous or non-MHC compatible cells, they can be administered using well known techniques which prevent the development of a host immune response against the introduced cells. For example, the cells may be introduced in an encapsulated form which, while allowing for an exchange of components with the immediate extracellular environment, does not allow the introduced cells to be recognized by the host immune system.

Transgenic and "knock-out" animals of the invention have uses which include, but are not limited to, animal model systems useful in elaborating the biological function

of polypeptides of the present invention, studying conditions and/or disorders associated with aberrant expression, and in screening for compounds effective in ameliorating such conditions and/or disorders.

All patents, patent publications, and other published references mentioned herein are hereby incorporated by reference in their entireties as if each had been individually and specifically incorporated by reference herein. While preferred illustrative embodiments of the present invention are described, one skilled in the art will appreciate that the present invention can be practiced by other than the described embodiments, which are presented for purposes of illustration only and not by way of limitation. The present invention is limited only by the claims that follow.

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising

(a) a nucleic acid molecule comprising a nucleic acid sequence that encodes an amino acid sequence of SEQ ID NO: 172 through 295;

(b) a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 171;

(c) a nucleic acid molecule that selectively hybridizes to the nucleic acid molecule of (a) or (b); or

(d) a nucleic acid molecule having at least 60% sequence identity to the nucleic acid molecule of (a) or (b).

2. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a cDNA.

3. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is genomic DNA.

4. The nucleic acid molecule according to claim 1, wherein the nucleic acid molecule is a mammalian nucleic acid molecule.

5. The nucleic acid molecule according to claim 4, wherein the nucleic acid molecule is a human nucleic acid molecule.

6. A method for determining the presence of a breast specific nucleic acid (BSNA) in a sample, comprising the steps of:

(a) contacting the sample with the nucleic acid molecule according to claim 1 under conditions in which the nucleic acid molecule will selectively hybridize to a breast specific nucleic acid; and

(b) detecting hybridization of the nucleic acid molecule to a BSNA in the sample, wherein the detection of the hybridization indicates the presence of a BSNA in the sample.

7. A vector comprising the nucleic acid molecule of claim 1.

8. A host cell comprising the vector according to claim 7.

9. A method for producing a polypeptide encoded by the nucleic acid molecule according to claim 1, comprising the steps of (a) providing a host cell comprising the nucleic acid molecule operably linked to one or more expression control sequences, and (b) incubating the host cell under conditions in which the polypeptide is produced.

10. A polypeptide encoded by the nucleic acid molecule according to claim 1.

11. An isolated polypeptide selected from the group consisting of:

(a) a polypeptide comprising an amino acid sequence with at least 60% sequence identity to of SEQ ID NO: 172 through 295; or

(b) a polypeptide comprising an amino acid sequence encoded by a nucleic acid molecule comprising a nucleic acid sequence of SEQ ID NO: 1 through 171.

12. An antibody or fragment thereof that specifically binds to the polypeptide according to claim 11.

13. A method for determining the presence of a breast specific protein in a sample, comprising the steps of:

(a) contacting the sample with the antibody according to claim 12 under conditions in which the antibody will selectively bind to the breast specific protein; and

(b) detecting binding of the antibody to a breast specific protein in the sample, wherein the detection of binding indicates the presence of a breast specific protein in the sample.

14. A method for diagnosing and monitoring the presence and metastases of breast cancer in a patient, comprising the steps of:

(a) determining an amount of the nucleic acid molecule of claim 1 or a polypeptide of claim 11 in a sample of a patient; and

(b) comparing the amount of the determined nucleic acid molecule or the polypeptide in the sample of the patient to the amount of the breast specific marker in a normal control; wherein a difference in the amount of the nucleic acid molecule or the

polypeptide in the sample compared to the amount of the nucleic acid molecule or the polypeptide in the normal control is associated with the presence of breast cancer.

15. A kit for detecting a risk of cancer or presence of cancer in a patient, said kit comprising a means for determining the presence the nucleic acid molecule of claim 1 or a polypeptide of claim 11 in a sample of a patient.

16. A method of treating a patient with breast cancer, comprising the step of administering a composition according to claim 12 to a patient in need thereof, wherein said administration induces an immune response against the breast cancer cell expressing the nucleic acid molecule or polypeptide.

17. A vaccine comprising the polypeptide or the nucleic acid encoding the polypeptide of claim 11.

SEQUENCE LISTING

<110> diaDexus, Inc.
Salceda, Susana
Macina, Roberto
Hu, Ping
Recipon, Herve
Karra, Kalpana
Cafferkey, Robert
Sun, Yongming
Liu, Chenghua

<120> Compositions and Methods Relating to Breast Specific Genes and Proteins

<130> DEX-0306

<150> 60/268,292
<151> 2001-02-13

<160> 295

<170> PatentIn version 3.1

<210> 1
<211> 591
<212> DNA
<213> Homo sapien

<400> 1
gctcttcctg tctacaaagg ggactgctca cagtggcctc agcttggtgg ttttgagggg 60

ccgccccccg gccctccata agggtatcct gggcctgaga attctgcatc tgccattgga 120

tggatgtaca gcctcaaatg gaagtgagtc ccacgggaga tgggtccgag gtccaggctg 180

tggccatcca gccccctgtg gcttgtccag cctctgtgca cccctggtgt cttcactcca 240

ggggcagaca gtagccactg cagttccttt cttcgtgaga taacagtagt gatagcagct 300

ggggctaaca ggctaggctt agtgtcctgc gcatttggtc agcttctcac tcgatcctcc 360

ctaaagcaat ggggaggccc ccactagccc agttttcagg aagtcaactg ggaggttaga 420

tgggggccag aggtcccaca gctactgatg gcccgagcca ggttgagctt tcctggatgt 480

ccagtccgga tcccacttgc agatctcatg ctctcagata ggtgggacaa gttcttttgt 540

cacagtgctg gctctgtcct gaggcctcat tgctggctgg tgtgctctgc t 591

<210> 2
<211> 2754
<212> DNA
<213> Homo sapien

<400> 2
gccagaagca gcctcagctt ggcaaggtgt ggagatgact gctgttccct tcgcatttgg 60

ggaaaacagg ctccctcggt agctcgatga tcctcttttg atcttgtgtg acctcctgga 120
gagtggatga cgctggtggc cttagctttt ctagacagtg taaattgcac tgggcgatgt 180
ccccagagca gggcaaggtc tctagagcgg gtctcccaca tgactggctt cacacaggca 240
cttccgctcg ggttgcatgc tctgtgtcat cttaccggtc cagggttgca ggtaggaaat 300
gtttgtaccc tcttctgatt gccacctcct tcccatcgcc ccttagggac agggcttgag 360
ggccagtgag gcgctggtca ggcaccccag gcctccttgg gacctgccca ggggcaccct 420
gagagctcct gaaaccccca cttagcttcc agacctttct gcaaaagctc ctcctggctt 480
tcctccctcc cccaatctat gggtcacagc taacagatct gagggcaact gctgtgctag 540
tggccagggc tgcacctgcc atccccggct ctgccacttt agggccttct agaggcagtg 600
tccttaggaa gtagctctga ggcatgggtt ttctgctcct gtgcagggca gctgatggga 660
taaggtgggg aaggacggtc agtgcttggg ccccagctgg ccagcctggc gatggggaaa 720
ccaaaccatg tccccccagcg aagggccaga gtgggaacct gtcctcatgc ccttcgtcct 780
gaggagccct gaggtgggca gcaggggcca ggggaagttt tcaggccttc atcaaagaga 840
acaacatcct cagctccgca cccctcatcc tgtatcagca cttaccggtg tgtgactgcc 900
cttgtcagct agcatacggt gggcccacct ggcccactgg ctgtttatgc cactgattta 960
tgatagggaa tattatcttt gaacccaatg aagtgttttc tcccccatca caaaaaaaaa 1020
aattcttatt tttagtagac atgtatttac caaaaatatg tactcaatta ttgtattttg 1080
gattttatca atttaaaaat tgtggaaatt tgtttgctct tacgccaaca taatattgat 1140
tttgcctctt ggctctgaaa gcccaaaata tttaccgtct agcccgttac agaaaaagtc 1200
tgctgactac tgagccagac ctccattacc tccatccctg ttggattatt taaagaaagc 1260
ctcagacagt aagggctttt ttaaaagaat aaaatgactt ggtttgcgct tggaagcagg 1320
ggaagcattc agatgagcgg tttctgcatt aaccctgcct atcacgcatc tcgtgtcctg 1380
tgtggctggc gagcccccct tggaaggttc tggtgcttca gctggctcct gcagagtcca 1440
ccccgcctcg tggtgggaat gcagagccct ttgctttcct tcttgccgcc tgcttcctgt 1500
tcctggggac ccgctgggcc tttggtctgc atcccctggc caggtccctc agggttgatg 1560
cgtggagaag gactttgagc agtggtgggc agcagtggcc tcctggccag ctcacactct 1620
tgtcctggga ggggcagcct gatctcacct ccacctagta ccttggggac tgaggacctt 1680
ttggcttctc tggagcctgc aagcctcttc ccatgtgtcc agctgctctt cctgctacaa 1740
aggggactgc tcacagtggc ctcagcttgg tggttttgag ggccgccccc ccggccctcc 1800
ataagggtat cctgggcctg agaattctgc atctgccatt ggaggatgga cagcctcaaa 1860

tggaaggagt cccacgggag atgggtccga ggtccggctg tggccatcca gccccctgtg 1920

gcttgtccag cctctgtgca cccctggtgt cttcactcca ggggcagaca gcagccactg 1980

cagttccttt cttcgtgagt aacagtagtg atagcagctg gggctaacag gctaggcttt 2040

gtgttctgcg catttggtca gcttctcact cgatcctccc taaagcaatg gggaggcccc 2100

cactagccca gttttcagga agtcaactgg gaggttagat gggggccagg gtcccacagc 2160

tactgatggc ccgagccagg ttgagcttcc tggtgtccag tccggatccc acttgcagat 2220

ctcatgctct cagataggtg ggacaagttc ttttgtcaca gtgctggctc tgtcctgagg 2280

cctcattgct ggctgggtgt gctctgctgg gaaaagcttt gcggggcttg cttggttaac 2340

cacagaagag aaggggactg tttggggtgc ctctctgcag cctccccgtg ctgggtggaa 2400

gcacggttac tgtgttctct aatgttcatg tatttaaaat gatttctttc taaagatgta 2460

acctccacac ctttctccag attgggtgac tcttttctaa aggtggtggg agtatctgtc 2520

ggggtggtgt ggcccttgga tgggtcaggt gggtgtgaga ggtcctgggg aggtgggcgt 2580

tgagctcaaa gttgtcctac tgccatgttt ttgtacctga aataaagcat attttgcact 2640

tgttactgta ccatagtgcg gacgagaagt ctgtatgtgg gatctgtgct tgggttagaa 2700

tgcaaataaa actcacattt gtaagaaaaa aaaaaaaaat aaaaagatgc ggcc 2754

<210> 3
<211> 856
<212> DNA
<213> Homo sapien

<400> 3
acgttaaaat taagaactta ggctttggtt taaaaaacaa taaatgaagt gaaaaaaaca 60

agccacagag taaaagaaga tacttgcagc aagtgataaa ggattagtat ccaggatata 120

taaagactgt tattgagtca atgtgaaagg gagaaaaaca cctgaagcaa agaatggatg 180

ccggcattaa ataggcactt caaagaggaa ccatgaacga ccaaaatcaa gtgagtaggt 240

gaccagttcc cattagtaat taggaaatag caaattaaga ccacaaagag ggcagtgagg 300

gtggctcaca cacctctaat ctcagcggct tgggagtcca ggcccagagg atcccttgag 360

gccaggaggt ggagtctagc ctgggaaaca tagcaagacc ctgtctctac aaaaaaataa 420

ataaataaaa taagaaaaaa gtaaaccaca aggagatgac ttaccaccag gcaaaaatat 480

taaagtatgc taataccaag tatcaagaag aatgaagcaa gatagctcaa atatgctttt 540

gaaggaaata tactgggctt ccattcattc tgaaataccc cttatttaag atactctatt 600

atattaaata cagttccaaa acaaaagaaa tccaaagaac aaaaaactaa cccaatactt 660

ttatcacttg taattgtata ttacaccata ttgaaagata tattttacga cttattagag 720

aacgattttt aaattggata tcactctgtg catacaaata aaataaagtg attaaggttc 780

taacaaaaaa acaaaccaca acaccaaagg ctttttaagg gggggaggaa taaggaaagg 840

ggcccaaaaa agggac 856

<210> 4
<211> 1580
<212> DNA
<213> Homo sapien

<400> 4
gtcccttttt tgggcccctt tccttattcc tcccccccctt aaaaagcctt tggtgttgtg 60

gtttgttttt ttgttagaac cttaatcact ttattttatt ttgtatgcac aagagtgata 120

tccaattaaa aatcgttctc taataagtct aaaatatatc tttcaatatg tgtaatatac 180

aaattacaag tgataaaagt attgggttag ttttttgttc tttgatttct tttgttttgg 240

aactgtattt aatataatag agtatcttaa ataaggggta tttcagaatg aatgaagccc 300

agtatatttc cttcaaaagc atatttgagc tatcttgctt cattcttctt gatacttggt 360

attagcatac tttaatattt ttgcctggtg gtaagtcatc tccttgtggt ttactttttt 420

cttattttat ttatttattt ttttgtagag acagggtctt gctatgtttc ccaggctaga 480

ctccacctcc tggcctcaag ggatcctctg ggcctggact cccaagccgc tgagattaga 540

ggtgtgtgag ccaccctcac tgccctcttt gtggtcttaa tttgctattt cctaattact 600

aatgggaact ggtcacctac tcacttgatt ttggtcgttc atggttcctc tttgaagtgc 660

ctatttaatg ccggcatcca ttctttgctt caggtgtttt tctccctttc acattgactc 720

aataacagtc tttatatatc ctggatacta atcctttatc acttgctgca agtatcttct 780

tttactctgt ggcttgtttt tttcacttca tttattgttt tttaaaccaa agcctaagtt 840

cttaatttta acgtacttga actgacattt tctaccctgg ccccctccca cccttagttc 900

ccagacacct cttatgatct ggggtcgagg agcccgcctt cctggcggtc tcagctgggg 960

cctggggagc gaaggcggcg ggcgctcgcg ggaggagctg cgcgattcgg atgctgggga 1020

ggtgaagctc gcggggccgc caggccgccg gggtaaggaa ggccgggagg ccgcgggggt 1080

ccacggcgcg gagggagccg caggcaccgg gcacagccct cgcccatcgc cgagacccgg 1140

caggcccagg agccagaggg cggcggcgtg agagggaacc gcctccaaag gacgccctcg 1200

ccctcccgca ggcatagtcg caggcgccag tcccggtccg agccagctgg gggtggctcc 1260

ggggagctga gccgggggag ggccgggccg cccaacggat caataggggg gtttctccca 1320

ggttgccgtt tctctggccc gcgacgccct acccgccgga gccgcccaac cgcaagcccc 1380

gccccgaagg cgggtgcagc caggaaggcg gggcctggtg gcctctgggc gctggcgcca 1440

agttcagagc cgcgccctgg gctgggcggt ggcggccgcg tctgcacttc ccccctgcg 1500

cgcctctgga gagcccggga gagacgcacc ctcaggtcgg ccaagaccga gaacaagcgg 1560

gcgcgggcag cggagcccat 1580

<210> 5
<211> 800
<212> DNA
<213> Homo sapien

<400> 5
tggtcgcggc cgaggtacaa aggctttgag gtccatggac tatacttgtc ccatttatca 60

tcccaggtgg tgctttgacc ctagggatac cctggctatt aagataaaaa gatttgtgga 120

cattaaaatt atgaatatgt cagtaataat ccagcacaca ttgaaatatt gacacagatt 180

accataattt gtgcaacatc ttataaacaa tgtcatttcc acagtagtct aaggcttcac 240

cagcctggcc cactgtatct agactttagg ttcattttaa ttaattatgc tttccttctc 300

tgtatcattt gggaagttga taaatatcac ttccttagat accttcattc agtgatatat 360

ctggctttta caattaaatt ggaaaaggta agtttctctt tggtgggttg agagttggac 420

catcaattct aatctacaaa aggaaattca tgatttcact ctgacgccta ggatctagcc 480

aaggctggtc tgcagtatca gatgtccaaa ctcatctact attagccata ttttgtgagt 540

cgtttgtcta aactttgtca aaaatgcctt tgccatgatt ttgttgctat ctggatttca 600

aacatggaca gttaggaaga tgtgcattga agtaggaaaa ttttgttcag catctgctgt 660

tatttatttt ttaccacttc aaaaatggcc actgtctttt taacaaacac caacgacaac 720

aacacacaaa acaaaaaaaa acaccctgcg gcttaccctg gccctccttt tccctgttga 780

attgtttccc ccccaatcac 800

<210> 6
<211> 956
<212> DNA
<213> Homo sapien

<400> 6
atttataagg cccttcaaat ttgtggcttc ctttctcata cttctcaagt ataatgaaag 60

ggggagaaaa accccaccat caacacaaaa gaaggctata aagactgtgc accttttaac 120

aagtcaattt gtagtcagtc cctgggcctg tctttttttt tttttaattt tgaagctacc 180

tgaggtttag aattccttca gccctagctg cttttattct gctttttatt taaacaaaaa 240

gagggggagg atctgaagga aactagtttt ctgtacaaag gctttgaggt ccatggacta 300

tacttgtccc atttatcatc ccaggtggtg ctttgaccct gccataccct ggctattaag 360

ataaaaagat ttgtggacat taaaattatg aatatgtcag taataatcca gcacacattg 420

aaatattgac acagattacc ataatttgtg caacatctta taaacaatgt catttccata 480

gtagtctaag gcttcaccag cctggcccac tgtatctaga ctttaggttc attttaataa 540

ttatgctttc cttctctgta tcatttggga agttgataaa tatcacttcc ttagatacct 600

tcattcagtg atatatctgg cttttacaat taaattggaa aaggtaagtt tctctttggt 660

gggttgagag ttggaccatc aattctaatc tacaaaagga aattcatgat ttcactctga 720

cgcctaggat ctagccaagg ctggtctgca gtatcagatg tccaaactca tctactatta 780

gccatatttt gtgagtcgtt tgtctaaact ttgtcaaaaa tgcctttgcc atgattttgt 840

tgctatctgg atttcaaaca tggacagtta ggaagatgtg cattgaagta ggaaaatttt 900

gttcagattt gctgttattt atttttaaa ttaaaaatgg aaatgtaaaa aaaaaa 956

<210> 7
<211> 489
<212> DNA
<213> Homo sapien

<400> 7
actatgtgtt aacataatcc caccttctta gagctttgtt ccttctgaag gtgtatagat 60

acagcttgtc ttgaaatgtc tttctccaca taatgaagca tgctgaatgc tgggaatctg 120

gagcagcagc cctgggagcc ctgagttttg aagtgttttg gtttgcttca aaggttagaa 180

gaacttgata tgtatggcaa acaactttag aatactagtt actcactaac atgaggcggg 240

taatgttgct ctagattcta tattccagta aagccagctt ttcttattat tggagtaggc 300

aaatgaatgg cattagaatt agtgggtggc ttgtaagttg tagttatagg cactttacca 360

cttcctgcca ttagcaggca tccttgtttt ttcttctttt ccctctttgt tccttctttt 420

ccctttctcc ttatacattt tctttctcta ctttaattct ccttcctcct tactgtagat 480

cccaagctt 489

<210> 8
<211> 3190
<212> DNA
<213> Homo sapien

<400> 8
ctctcattag cctgttcaga gtcttggggg aaattgagat ttttgagatt ttttttaaaa 60

actcaaatat tttactagtt tgcctgccat tttatttctt ttacaaagca gaagcatata 120
ccaatttatc acagtatttt agtaaatact gcaacattca tccttaaatg ttcaccaaga 180
aaagcatctt tgtagtagtg ctggaaaact attcagaata tacagataaa aatgctgttc 240
tttaattgct tacattgctt cttcccataa aaagcaaaaa ggaatcagtg cttgctattg 300
ctcctttcct tgaagttgta acaattgata catatattat gagttgactg gtcgattctg 360
tacctggccc atcctttaga atgttcttgt catgtagcag tcctacgtac tcttttcatg 420
agcagtctgt gatctcactc tgtgagttca gctattactc gctcgtggga gcttaatctt 480
ttcaaaatga agttgattta aaaagtcttc aggcagagta atcatgttag aggtggtatt 540
cgatggaaga aagtttagag agttaggagt gggggtagaa ttctagaatt tataagagtc 600
caggaagcat agcagtcagg ggcaaaaatt agcgtaatat ggagtaggca atagaggagc 660
tactggagtc agaagtcact gcagagtgca acataggaag atggactcct agcttacatg 720
agattccctg cagctgtaat atagacaatt cccacatggc tgttctacac agaattacct 780
gctaagattt tttgtttatt tttgtttgag tggtattttc actccaattg tataatggaa 840
atcagtggga aaatagggtt taccttatat tcatgagttc tagtttctac tgttctgcta 900
tgtgtttcta agcaagagca aaggatactt catacttttt tcgttatatg attgatcttc 960
aaattgggat ttaccttttt caatatgttt taaagtagtc ttattcctct tttgatttgt 1020
taaacaagca ttttagttca gctattgaat agccttccaa aaaattaatt cagccttgca 1080
ggtaagtacc atactaagac tttaacccaa tagtttttaa tcattctgcc tttattccaa 1140
actgtaaatc tgtacacata agataaaaca tactaagtat tgcataaatt gttaacgtta 1200
cagtaaattg ttatctgcag ggctgacaga cataatgttg gtgggcaact gtgatcctat 1260
acatacatat atgcaaaagg ggattttaaa agtgcagatt atagagtaga ttgacaaatt 1320
ttattttata ttcagttgtc ctctctgctt ccatctgtgt tgctctctta gttgagagag 1380
agttagccat ttgacgattt taagtcagtg ggaacttatt tttagttact caataaaatt 1440
aatattttat ttgtatttta acttacagag taggttggta ataacagctg aactgtgtaa 1500
cattgttgct tcaaattgaa gtttatatta tgaacattca gaatcaatgc tcatgtagca 1560
gcatattatt gagctatttt gagtttgaaa tgtggagaaa cgctaaacca tgtactatgt 1620
gttaacataa tcccaccttc ttagagcttt gttccttctg aaggtgtata gatacagctt 1680
gtcttgaaat gtctttctcc acataatgaa gcatgctgaa tgctgggaat ctggagcagc 1740
agccctggga gccctgagtt ttgaagtgtt ttggtttgct tcaaaggtta gaagaacttg 1800

atatgtatgg caaacaactt tagaatacta gttactcact aacatgaggc gggtaatgtt 1860

gctctagatt ctatattcca gtaaagccag cttttcttat tattggagta ggcaaatgaa 1920

tggcattaga attagtgggt ggcttgtaag ttgtagttat aggcacttta ccacttcctg 1980

ccattagcag gcatccttgt tttttcttct tttccctctt tgttccttct tttccctttc 2040

tccttataca ttttctttct ctactttaat tctccttcct ccttactgta gatcccaagc 2100

ttctagctta ggtttgcaag tcatattgct tggccctcca cattcactga gaggtgaaga 2160

taggctgacc ccctgtcctc ttacatttga gggatcatag actgctgtgt gaattctgga 2220

aagtctcagg tccctaccag ggcactgaat ggcttctcaa tggctgtaga gacagtacag 2280

ttttccaaag cagcctaatt catctggaca gctaccaggc actttggaaa gttggttcag 2340

ttactactat gaggccataa tatatttgct ggtattaaaa ttcttcagaa ttggaattac 2400

tatttgaaat aatattttgg ttgacttaag ttttgagaga caattctaaa attgatctag 2460

agactcattc aatagcaatg tgacctttta aatacttaca ttaagtaaaa ctgccagtag 2520

attaaatcat atatatatat atatatatat atatatatgt aagagcttcc tctatttact 2580

actgttgaac ttcagtaatt tttagaggct aaataatggt cagaatgttt ttaagtgtgc 2640

tcttttatta catgcttgtg caggttttgt aattcagtac agaaaagttt aaccttgtac 2700

atttttgtat gtaaaaagtc ttttaagtag tcttatcctt atttaaataa acagaataaa 2760

attaccttga gtaggtctgt tattcttatt aaaatggaaa aatgctctgt aatgacttga 2820

tctgttttta tttgagtgaa caattttgga aagtattctt tatagtacaa ctttctatac 2880

ctggattgat taagatcaga tgtgattcga gtagtccagc catatcttgt agcccttctt 2940

tgaatgagag ggtggctgga gtggtctggt gctgggatat cacggtgcta cagagcctga 3000

catgttgact gtcactacat gttgagggat ggaaatagaa gtctctgaac ttcccatgta 3060

atattaaagc tcttaacaaa atgagacaaa ctagagattc agttgagaga ttttatgtta 3120

gagtgatctg aaaaaaagtt aatttctaaa ctgctatctt aatattatta tatttggaga 3180

ctgatgctgt 3190

<210> 9
<211> 672
<212> DNA
<213> Homo sapien

<400> 9
ggtcgcggcc gaggtactat tgctctggct cctggccctc tccttgctat gggtcttacc 60

ctcaagtcgc tctgtgattc aaagatgaac tgccaatcaa atgttcctct aatgaaagat 120

ccaatcactc tacagcatgt gtgtattcaa agaacctatc taagactttc ttttggtcat 180

ggtgggaggc tgttgctgaa aacataccag agcccattgt ggaggtcagc tgacaggccg 240

catgaccttg gcaatggact actggtcatc tgggactgct taggactgtg caatggaact 300

tgggggcaaa actgatggag acagccaatg ggccttaaat ccagcaggca aagacagagt 360

aagttcttat ttgtgtagcc cagggcttat caaagtgtgg ttcttggacc acgtgcatca 420

gtatcagctg taagtatttg gcaaaatgca gattcccggg ccctgcacca aacagattga 480

ctttgaatct ctgggggttg ggctaaaaaa aaagaaaaaa aaaccctaca ttttaaacaa 540

gctcttcaga tgacccttgt gtaagtttga gagcatctgc tggaaaacca ctagaatttg 600

caaacggcac tcaaaatact ccagccagtc cactagccaa agaccagatc tgagaccgga 660

tgggaaatta tc 672

<210> 10
<211> 997
<212> DNA
<213> Homo sapien

<400> 10
ggtcgcggcc gaggtactat tgctctggct cctggccctc tccttgctat gggtcttacc 60

ctcaagtcgc tctgtgattc aaagatgaac tgccaatcaa atgttcctct aatgaaagat 120

ccaatcactc tacagcatgt gtgtattcaa agaacctatc taagactttc ttttggtcat 180

ggtgggaggc tgttgctgaa aacataccag agcccattgt ggaggtcagc tgacaggccg 240

catgaccttg gcaatggact actggtcatc tgggactgct taggactgtg caatggaact 300

tgggggcaaa actgatggag acagccaatg ggccttaaat ccagcaggca aagacagagt 360

aagttcttat ttgtgtagcc cagggcttat caaagtgtgg ttcttggacc acgtgcatca 420

gtatcagctg taagtatttg gcaaaatgca gattcccggg ccctgcacca aacagattga 480

ctttgaatct ctgggggttg ggctaaaaaa aaagaaaaaa aaaccctaca ttttaaacaa 540

gctcttcaga tgacccttgt gtaagtttga gagcatctgc tggaaaacca ctagaatttg 600

caaacggcca cctcaaaata ctccagccag tcccactaag ccaaagactt tcttttggtc 660

atggtgggag gctgttgctg aaaacatacc agagcccatt gtggaggtca gctgacaggc 720

cgcatgacct tggcaatgga ctactggtca tctgggactg cttaggactg tgcaatggaa 780

cttgggggca aaactgatgg agacagccaa tgggccttaa atccagcagg caaagacaga 840

gtaagttctt atttgtgtag cccagggctt atcaaagtgt ggttcttgga ccacgtgcat 900

cagtatcagc tgtaagtatt tggcaaaatg cagattcccg ggccctgcac caaacagatt 960

gactttgaat ctctgggggt tgggctaaaa aaaaaaa 997

<210> 11
<211> 696
<212> DNA
<213> Homo sapien

<400> 11
gccgcccggg caggtacaaa tggtgcccat gccattcatt tgactgtggg tggccctcta 60

gtctagggct ctcttagtga atggttgtgg aaatatgatt tttctaagtt ccttcctttt 120

ccttttgata gatgagtttg agatgatgga gtaggagtga ggccctcagg cacttctggt 180

aaagacattc cacctgcaag cagcattttg agtaaagcac tgctgtggtt tgccgattta 240

tggtccattt aatgttaggc taaagcacct ttaatcattt ttgttgtttt aagataatgt 300

atttgtgaag tggataaaca ctggaaatag ggtgcttctt ctggaaagtt cagtgtaaaa 360

cactaacaag gctttggcgg gtttatctgg ctttataaac aagtctgaaa aatggatgaa 420

agctaaatat ataaagcagt tggttgtcta tcttttatca tttttactc agatctgtat 480

ttaacactta tttatttgtt agtttttaca ttcaaaagaa actacacttg gaactttggc 540

taacattgta ggatattttt taattgttcc tacattttta agcatgattc atcattttgg 600

taacttagat catttttaat ggtcttttct ttcaataacc agttacatca tgttttggga 660

actctttggt tccatataag gtgaattggt gcaaaa 696

<210> 12
<211> 3233
<212> DNA
<213> Homo sapien

<400> 12
aacggtccta aggtagcgag agaatactac caggtgctag tttttccagt attgacttct 60

gattactatt tccttttctc atctttagtt tttcaagatt tgctttacca aaatagtaaa 120

gcctttatca tcagcttata ttgaataatg ttgtaattgg tttcaatcaa agtttctcct 180

caggtacttg ggggccccta gccttctaag gaactcccag gcacctactt aacaaggcca 240

gctacacact cagtatgtga taagccccat gatggatgca ggttagaatt caaagacctg 300

gttggagtcc tagatgtgga gacaggatca tcaggtcaca cttgttagat gactaacact 360

atcagtagaa gctcttgaga gattttccta acgcagcaag atttctgtga gtagaggtat 420

cctgggaggt atcctgggag gcagcctatt gacttgacca agtaagctga tcaggtggcc 480

tcctctaccc actaaagaaa tgtgtaaaca ctagcaataa ttgctttatc ttaaactcct 540

ggacatactc agttcctcca ttccactgtt ctattgccaa tacctttgtt gtttcttca 600

cactcctctt ggcagcaaat gtctgaaagt atttcaattg tgtaatgtta aggagttttt 660

tcatagcttc agaaaagagg gcagcaaata tgaagcctta agttcaaaat aagtcattct 720

acctagaaat acagacccca gagcacattg catgaaaata cctgtactct gcagttcctc 780

aaagcagtat tcttcctgaa aagccaaaca ccacacctat tttcctattt gctaagaatc 840

agaataagca cgttgtaaat agtatccaaa gcagattcta aaatgacata gtaagaagcc 900

agattcaaat tgtaaccaaa gaagacaata gaaatcccac tttaccccac tgtcatcagt 960

tagaacaccc ttgcaaaaac tgtaaccact taagcaattc atctgatccc agaagatcat 1020

accttctttt gaaagtatag gacagatatc agtgggaaac gtcggcgttc tgagcaacac 1080

aggataaatg taggagggcc ttaaaaaata aatctcaatt catacactgg agcagcaaaa 1140

aactgagcag gaaaggaaac agaatccaaa gtcatttttc atatagctgt tgtcaaatag 1200

tataaccttg gtgtcttctt tgagttgcct ggacagtatt tatgaaacaa aaaactaaat 1260

gccccccatt tggggacggg gggaggggtt cagacctcta acctggattc agagccttag 1320

aggccgagag ggaatctgga atctggtatt actgagatcc taggtaaaag aaccagcctg 1380

gcagtctttc ccacctcatt ggtccgtgct tttattttta aacccaaaaa aaaaaaaaaa 1440

acacaccctc ttatgtagga atttcccttt tacaaataat ttgacctggt agaaataaac 1500

ttgcctgcct gctcttaaat gccagacagt tggaagcaaa tgccgaggga aaggtgccca 1560

gagccatgct tgataggact ttgaatattt tctccttaat taaagtacgt tgcttgtatt 1620

tagactataa gatgctatgg aggtctatcc catgccatgc caatgtgaat tgctttgctt 1680

cagtaacaat cagaagacca gtccaacaga aaataacttg tcataattcc accttagatt 1740

ctagacctct catacctgca gtgtacagaa tatgtacatg ttccaatgga attcactatt 1800

tttggcttta gtgtcaaaga gattggttct acaaggttca tctgatttcc cataacaagt 1860

aaattttata atcctatgat tctaaattca atccccaata tagattctaa gcatcaaatc 1920

aaaatcacag acaaagggga actggtcgag aggggtctta gttatttcaa atccatgacc 1980

aaagtgtcca aagacatgaa actcttatac ctgctgagca tttcacttta ctatacaaaa 2040

tgtcagctac ccagttgcat cctgtgacat gatcagactg tcaatgtgga ccagtggcca 2100

ggagcatatt tatgggccat ttctgttcat cattctttac agagcattga ggtttcccac 2160

tgaaacagct tctttagtca gacgtctata gattttacat aaatttacat ttaaatgcat 2220

taagttagat ggcccaattg agcatctgaa tgaatatagt gggggttggt ggtggtgcaa 2280

attctgctgg ctttatgtta tggttttctt cgtgtttttt cttggttttg tctggcttct 2340

tctggcaagt gccctaaaag actggaacac tgtataaagt catagacata gaaccatatg 2400

ggaaagccca gatgaaaaaa tggaagaata aaatcaagtt gtcaaagttc cagcaacagc 2460

cctgacttct tcaggaatcc aagcaaattg aaagccaaga caaaatgtac aaatggtgcc 2520

catgccattc atttgactgt gggtggccct ctagtctagg gctctcttag tgaatggttg 2580

tggaaatatg atttttctaa gttccttcct tttccttttg atagatgagt ttgagatgat 2640

ggagtaggag tggggccctc aggcacttct ggtaaagaca ttccacctgc aagcagcatt 2700

ttgagtaaag cactgctgtg gtttgccgat ttatggtcca tttaatgtta ggctaaagca 2760

cctttaatca tttttgttgt tttaagataa tgtatttgtg aagtggataa acactggaaa 2820

tagggtgctt cttctggaaa gttcagtgta aaacacaaac aaggctttgg cgggtttatc 2880

tggctttata aacaagtctg aaaaatggat gaaagctaaa tatataaagc agttggttgt 2940

ctatctttta tcatttttac tcagatctgt atttaacact tatttatttg ttagttttta 3000

cattcaaaag aaatacactt tgaactttgg ctaacattgt aggatatttt ttaattgttt 3060

ctacattttt aaagcatgat tcatcatttt tgtaaactta gatcattttt taattgtctt 3120

ttcttttcca atagaccagt taccactcat gtgtctgcag aacctcttta ttgtattcct 3180

ataataaatg taaaatattt gtagcaaaaa aaaaaaaaaa aaaaaactcg gtc 3233

<210> 13
<211> 847
<212> DNA
<213> Homo sapien

<400> 13
actagactat gatatggact taccctcttg gtgccagccc tacaagctgc atgaccgtaa 60

tcagcctgtg acactacgac atgcgccact cagcctgtgc cactcttggc ctgagccttc 120

ggcctcttat gactgaggcg gacaactcac gctcaacaat ggcaaagact gctgcaccat 180

tgctagatca catcaatggt gccaccaact actctcttct tcacctacca acagtggact 240

gactggcttc tatgactcct atccacctca tctgctcccc tagtcacgaa ctacaagaca 300

ccacacaccc ccagccgcag cgcgaatgcc aaaggttcag cacacacggt gcgcaaacaa 360

cccaatgcgc gacccatcat catccataca tatctggcgc tgctacgcgg acctacctac 420

gccatgtcgc tcctgactac tctgctccac tgatggctcc tcctaccaac acgcgcttgg 480

ctcctgccag cctgcagccc acacacctgc gccctccctt ggcacgacac ccactcaccg 540

ccgactgccg aacgcaccag ctgacggaca cacgtccact ccacccgcgc ccaatcacct 600

cgcgcacgcc ccagcccttg ccctcccata cccacggcct acaccacacg cgcccacctc 660

acactgcgac cggctgccca tatctctcca cgtcccgccc tctgcccccc ctccacacac 720

gcagcatcca cccagacaac ccacactgca cgacccctca tcacagcccc tcaaagccct 780

ccaccaccac acaccagcag tcccccgccc caacacctaa taaaccccac ccccgccgag 840

cctcaca 847

<210> 14
<211> 267
<212> DNA
<213> Homo sapien

<400> 14
actgtagcag tgagctcaag tgttgggtgt atcagctcaa aacaccatgt gatgccaatc 60

atctccacag gagcaatttg tttaccaaga atctaagaat taaatcttag aatgtattaa 120

tgttaaattt ctgtgagatt atattgtagt cacgtagaat gtcctgactt gtaggaatac 180

ccactaagga aatcagaaat cacggtagag cgtcagcaat ttactctcaa atggttcaga 240

gaaagaaagt tctttgtagt aaagctt 267

<210> 15
<211> 824
<212> DNA
<213> Homo sapien

<400> 15
tggtcgcggc cgaggtacag tgggtggaaa gggcatttgg agctcattag aatgagacat 60

agttaagagt cccattctca tcagtgtatt ccagactgag gaagaaatgg ggcagcagtc 120

aggagagctc gggattttga gtatagcaga atttaagtga aatggaaact acactcttta 180

atttgttctt ccatggaatt gctttttcta tgcaagggct gagcccccag gagagccctt 240

gtgaagggaa tgctgatttg tgtgaatatc tgtaggtgag taggtatcta gtgaggatga 300

gttgggagga tgagttgggt aaggcgtgcc cctctgacac tgttctgggt ataaaagaca 360

acatcatgat gagatcttca tctgaataaa actatgccct ggccttttca gaaactgcgg 420

gcactgcagg tcccacagtg tgatggagtc caagctggga tcactgcgag atgaggagtc 480

agaggagtgg cttcggcagg catgggagct tcaggccctg agagagaaga cagaaattca 540

gaaaacggag tggaaaagaa aaacgtgaag gaactgcatg aagagcacat ggctgagaag 600

aaagagctac aggaggagaa ccagaggctc cagggcctcc ctgtctcagg atcagaagaa 660

ggcaggctgc cagttccaag tgccagatca agcaccctcc gtgccagctg caggaacgag 720

ctaggatcat tgcttccagg aggagagacc agccttggtc tcaaggaagg gcaccggacc 780

aaaggggcaa ggggggaca cagagaggat ccacaggaaa aatg 824

<210> 16
<211> 1998
<212> DNA
<213> Homo sapien

<400> 16
tttactttta ttaaagtata ggaatcaaac tggataccaa attctcagtg cagttgggta 60

gtcattttgt taatgtattt ttaaaaaatt ttaagggtaa aaaccagcaa gattccattt 120

agaatgattg tgaaaaaaac actgtaagac gtccattttc aaaatgcaaa aaatgattct 180

tcctgatgtt aggaaggcca atgaaaacta tatgtatatt gaaaatattt tttcctcaaa 240

actttttccc tgatacagaa gtctgagagc ttactttggc tacattacct gactaaagag 300

agaactttag attagacctg gggtaaattg agatgccaag ggagtgtcta gctaaatgga 360

aataccacga aggtttgtaa tgccaagaaa gtcagctctg tggtgtgtca taagcagcat 420

atggaaacca ggagtgacac attagaaccc gggagttgtg catacatctg atcaagcatt 480

tgactctgaa aatattcagg gagtttagaa attgttaacc tttggaacca gtattgttta 540

gcaatagttg agaagtgtta gcaagaatga tatcaagtta aacttaggca cttggagtta 600

catccttaaa gccttaatag ggcttatgag ttttatacag tcatacagat agaaatatgt 660

tgcttttgtt actacgacag tcatatatta taagaaataa tcaaaggtgg gtggaaaggc 720

atcctctctt tgatccaatt ttctgtacct ttttcttcag gtcacacaca ctgctagccc 780

aggaatcact aggtattgat gactctactt caagctgtgc aaagcccttt ctggagacag 840

ccaggatgtt ttgtagggag agaggcagga gtcctcaggg agtggcctgg ggtgagaccc 900

tcccataggc tctaagagtc tcattctcat cagtgtattc cagactgagg aagaaatggg 960

gcagcagtca ggagagctcg ggatttatga gtatagcaga atttaagtga aatggaaact 1020

acactcttta atttgttctt ccatggaatt gctttttcta tgcaagggct gagcccccag 1080

gagagccctt gtagaaggga atgctgattt gtgtgaatat ctgtaggtga gtaggtatct 1140

agtgaggatg agttgggagg atgagttggg taaggcgtgc cctctgaca ctgattctgg 1200

gataataaaa gacaacatca tgatgagatc ttcatcatga aataaaacta tgccctggcc 1260

ttttcagaaa ctgcgggcac tgcaggtccc acagtgtgat ggagtccaag ctgggatcac 1320

tgcgagatga ggagtcagag gagtggcttc ggcaggcatg ggagcttcag gccctgagag 1380

agaagacaga aattcagaaa acggagtgga aaagaaaaac gtgaaggaac tgcatgaaga 1440

gcacatggct gagaagaaag agctacagga ggagaaccag aggctccagg gcctccctgt 1500

ctcaggatca gaagaaggca ggctgcccag tcccagtgcc agatcagcac cctccgtgcc 1560

cagctgcagg aacaagctag gatcattgcc tcccaggagg agatgatcca gtccttgtct 1620

ctcaggaagg tggaagggat ccacaaggtg ccaaaggctg tggacacaga ggaggactct 1680

ccagaggaag agatggagga ctcccaggat gaacagcaca aggtgctggc agctctgagg 1740

cgtaacccca ctttgctgaa gcacttcaga ccaatcctgg aggacaccct ggaagagaag 1800

ctcgaaagca tggggataag gaaggatgca aagggaatct cgattcagac tctcagacac 1860

ctggaatccc tgctgagagt ccagcgggag cagaaggccc ggaagttttc tgaatttctg 1920

agtctgaggg gaaagcttgt caaggaagtc accagcagag cgaaggagag acaggagaat 1980

ggcgctgtgg tgtcccag 1998

<210> 17
<211> 653
<212> DNA
<213> Homo sapien

<400> 17
gcgtggtcgc cggcgaggta catggccgca agcagactaa cgcgttgacg ctaatttaat 60

gtattttacc tcacactaag gtcatgcttg ataaagacgt taaactcaac ttgtaaaatg 120

gtagcccagt gctatgcaca gagtgggtgc tcattagtgt tgaatgaaca catttgtaat 180

actacatgta attccatctg actgctttgt taaattttca gttagaacgt agatactgta 240

aagtccacac acacattaaa tcttgttttc ctgaaagtat ggcatcaaaa atacttgtag 300

aaaaaccttg tcacaactga tttgaatgtt cctattttct ttgactttga tattggcttg 360

taatgtctct tttcatcata tgtaatatca gtggaacagg cagcgctact caagtcctaa 420

ggattcctca gtgatcagtg atccagggcc gttcatgaac cactgggctg gatttgactg 480

ttgagtgtgg cagttaatgc ccctcaagaa atcaaaggat gtcttataag tgtcttccaa 540

aaaaaagcaa atgctgaaat cctattggca aagtaaactg aaattggctg ctatatttta 600

tataatcatt tctgcaaatc ccatttttg aatactaata tttgacatgg tta 653

<210> 18
<211> 1498
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (29)..(29)
<223> a, c, g or t

<400> 18

ttattcagtg catagcttta agccagtgnt ggattcacta agtggacagc cagtctccca 60

gctctctgcc ttccccaaaa gggtcgtagt aggtcacccct tctacagcag ctaactagag 120

tcctaactaa tgggatccag cagggccatt tctccagagg gccagtatcc tattaggaga 180

ctcttggaat tcttaggttc tactcaagag tggaaggacc aatcacctct gatattctgt 240

ggaaggtttg gggtcaaatt ctgccctctg cattctgtgc aacttgtata aaagtcaagt 300

tagtattaca tgaattgggg tagggttagt gctttgaaaa aatgttgaac cggctgggcg 360

cggtggctca cgtctgtaat cccagcactt tgggaggccg aggcgggtgg atcatgaggt 420

caggagttcg agaccagcct ggccaacata tagttgcttt ggacctcatt tggaaaaata 480

atctgccttt ctaattgttc tgcataggtt aaaatgataa atttacattc tttgaaccta 540

taccagattg tggtgtccga gtgaccggca cactgtctga cacacagtca gtgtgcacgt 600

atttgtctga gtgaatgagg agacctgaga aaccggtgac gtggcacagg gaagccagct 660

ggcccaggat tccgtacatg gccgcaagca gactaacgcg ttgacgctaa tttaatgtat 720

tttacctcac actaaggtca tgcttgataa agacgttaaa ctcaacttgt aaaatggtag 780

cccagtgcta tgccaggagt gggtgctcat tagtgttgaa tgaacacatt tgtaatacta 840

catgtaattc catctgactg ctttgttaaa ttttcagtta gaacgtagat actgtaaagt 900

ccacacacac attaaatctt gttttcctga aagtatggca tcaaaaatac ttgtagaaaa 960

accttgtcac aactgatttg aatgttccta ttttctttga cttagatatt ggcttgtaat 1020

gtctcttttc atcatatgta atatcagtgg aacaggcagc gctactcaag tcctaaggat 1080

tcctcagtga tcagtgatcc agggccgttc atgaaccact gggctggatt tgactgttga 1140

gtgtggcagt taatgcccct caagaaatca aaggatgtct tataagtgtc ttccaaaaaa 1200

aagcaaatgc tgaaatccta ttggcaaagt aaactgaaat tggctgctat attttatata 1260

atcatttctg caaatcccat tttttgaata ctaatatttg acatggttaa ttcttattaa 1320

tttgttggaa ttgtttattg ttaataatgc aaatagataa tttttaatta tccacaactg 1380

atttgaatgt tcctattttc tttgactttg atattggctt gtaatgtctc ttttcatcat 1440

atgtaatatc agtggaacag gcagcgctac tcaagtccta aggattcctc agtgatca 1498

<210> 19
<211> 171
<212> DNA
<213> Homo sapien

<400> 19
gccgcccggg caggtactaa atgaaacata atatttattt ataaaagtgt gtagattgtt 60

aaatcacaaa aagagtgcta tgaccattat gtatgaggaa acaggccttt gacctcctgg 120

aaagcactgc tcaaaagtca ttagtgccca tttttgaatt ccccaaacag a 171

<210> 20
<211> 1820
<212> DNA
<213> Homo sapien

<400> 20
gaatttcgta atccttgaaa ttgaaaaaaa aaaaattgtg tttttaaaga gtgaaaacag 60

ttaggaaaca agtagaactg taatcagaac gctgcttcaa ttgatattaa aaataacctc 120

aataataatg taaaggttcc tttctcttgt gtcagttata ttcttaggga tagcctagaa 180

ggaatatatg gttagaacta agtgtgacta atcatctgag ccttgaagag aaacttcagt 240

gcctctaaac agatcatcta caaaacaaca ggtaaacatt tatgccagtt aagtgggtca 300

tgtttttgtt tcttgggttt ttcctaaatt taagtgaggt tgggcttacc ttgtagataa 360

aattatgttt tctttttggt aaatacttga atgtggataa cgtcaaatca gaatattttg 420

tgaggaggtg atgatttgaa attaagctag atttctaggg aggtgttggt tccaatgaag 480

gatgggaaga aattaaaata gtcttcaaac ttcttcctta ttatatttgg ttgctttgga 540

aaagattggt cctatcctca atctaattta ttcactatta atattttaaa aacattcctg 600

agatacttaa aaagacccac ttagcgatta tagttgctca atgaaacaag aatttattta 660

tgcatagatt tttctctgta tcttaccaaa atccacttta cttagataac actaaattgt 720

tcttaaagac tactcatttc ccaataatcc tttatgattt caaaatttct agtggctcag 780

aagtgaattt tattttattt gtctttcact tgaataaatg agaacccaga aattaataat 840

gttgtttatt gcttactgtc aggactattt caaagactaa gaagagtttc ttctaacccc 900

tccctctcaa aggaatccta aattattagt tgttagataa gttttgtatg ctaagatatt 960

caggtttata gtttatgtat gtgtgtatat atataaatat atatgtatat ataaatatta 1020

tgttcagttt ggagtctggc acaactccat tatgtggatt agagagtaag atattatgga 1080

tgataaagta ctaaatgaaa cataatattt atttataaaa gtgtgtagat tgttaaatca 1140

caaaaagagt gctatgacca ttatgtatga ggaaacaggc ctttgacctc ctggaaagca 1200

ctgctcaaaa gtcattagtg cccatttttg aattccccaa acagaaagct tcttagaaaa 1260

cacgctgaga ttttatttac agggaattct ttgacacatt tcaattggtg tgtagtcaag 1320

tatagcaagt acttaataat gactgaattt catgttccta cagtcataca tattcattag 1380

aagttttatg ttgttggtct gatctgattc ttctttgttt gtgggtggaa cggcactgag 1440

agaagtatag ttttttaaac ttgaacatgt tcagtagtta cattgcctta gaaaacccag 1500

acacatagca gtggaaatga aagaaatggc atcagaagtg acttaattta gcaattgtga 1560

ttcctcttgt aaaacaaaac aaaaaaacaa tgccatattt tttggagaaa agttggcaat 1620

ataggggttt cgttgtctgt ttcacaagaa gactcatttg ttcttttggg ggaaccagtg 1680

ccttacagat tttgtatata ctgtaattat tcaggactag ggaacaaaca attgtattgt 1740

atttgttaca gattgtatat ggctttgttt taacattccc ctaaataaaa tggcttcatt 1800

ctccccttgg aaaaaaaaaa 1820

<210> 21
<211> 611
<212> DNA
<213> Homo sapien

<400> 21
acccaagaca ggttctgaac catgcttatg cagagctttt agtattaaag agggagagta 60

aaagaagtgt cagagtccag atttatcact gaacccaata ctttcttact ccctggggca 120

tctcctaata ctgatcctaa aatgctcctg tttctgagaa gctagggcaa gacctgcctt 180

acaaagacca gccatttgcc ttattcatag gatcataagc aagagaactg cattccagga 240

agaatgaagg aagaaggaag gctgctcaca gtagcagaag ggaggcaggg gccgagctgt 300

tcaagtcaca taaactctaa gaagcccagt cagcaaaata agtctatctt caattctagt 360

tgagtccagg actctgagga gctgtgattc acccagtttt tcctgcaaaa ggcacagtcg 420

ctaaactaaa ttggtgcaat tcacttcctc ttgcctctct ggttcattcc accaattgtg 480

gttgagaaac acatcttagg gaagaaacag tatctaagca ttaaagagaa aatatcccac 540

tttgctcctc ttcctcccta aacccgaact gctcttacat acaagataat ttttaaatca 600

taagattggt a 611

<210> 22
<211> 1885
<212> DNA
<213> Homo sapien

<400> 22
catgaacatt tgaggctgat tccctgtggg aaaaatcatt caaatctatt cactcatctg 60

atggctgttg cttgttttat tttttgtcca agagaggtgg tgttggaccg aggtagagaa 120

gacagtggta caccagaaat aacccaaagg attgcccctt ctgtagaagg cccttagact 180

ccatgatgcc tttcagctgg gtgctatact tgcacctaac tctgggggct tcactttcta 240

tccctacaat tactcaaaca gataaaaggc tggatgttaa catgtagtta taaggggcgt 300

gatctaatag taaggaatat cacttcccac aagtccttca aacaagattt gtgaggagct 360

ggatttgtca gcatgtcaga tctttttgaa aaccagagag tagaatgtaa gcaataccct 420

tgtcgtaatt aaagaccaga ctccatcctt ataccactga tgcctctggt accttaatcc 480

ttaaaatatt tagtgaccct tgccttctaa ttcttgacac aaatatataa tgaccatttt 540

agatcgggga actccctttc tttgaaggca gtttagggat tccacagatg ggctttgaac 600

ctgctaaatg tgtatggaaa actgagtgaa ttacaaatgt ctttttctca aaagtgcgtt 660

tctggtttct gtcagattca acaggtctgt acccaagaca ggttctgaac cactgcttat 720

gcagagcttt tagtattaaa gagggagagt aaaagaagtg tcagagtcca gatttatcac 780

tgaacccaat actttcttac tccctggggc atctcctaat actgatccta aaatgctcct 840

gtttctgaga agctagggca agacctgcct tacaaagact agccattttg ccttattcat 900

aggatcataa gcaagagaac tgcattccag gaagaatgaa ggaagaagga aggctgctca 960

cagtagcaga agggaggcag gggccaagct gttcaagtca cataaactct aagaagccca 1020

gtcagcaaaa taagtctatc ttcaattcta gttgagtcca ggactctgag gagctgtgat 1080

tcacccagtt tttcctgcaa aaggcacagt cgctaaacta aattggtgca attcacttcc 1140

tcttgcctct ctggttcatt ccaccaattg tggttgagaa acacatctta gggaagaaac 1200

agtatctaag cattaaagag aaaatatccc actttgctcc tcttcctccc taaccccgaa 1260

ctgctcttac atacaagata atttttaaat tataagattg gtattaacac aattattgat 1320

aaagagaaac aatgaccaac tcattagcta acgatgctag aatacttatg caagccctag 1380

agttaagggt cttagtgtgg acacctttcc agaattggaa ggaaaaccaa ccagaaagct 1440

tattaccctg catcagctga aaagctaagc cacagccatt ttccctaaag ttctgtttct 1500

gggagaatga gatcttcaag aataactctt gccccttgat gaggcagtca aattcaaacc 1560

agtgatggca acaacttgca aacacgtaat tcctgcccta attttccagc acttaaaaca 1620

aaatccccac tcaatacaaa gtttctatgt gcctcttgcc tgaaatcaac aagaaacagc 1680

tcacctgccc aaagactcct ctttctctgc cagggcaaaa gcaatctgca gcccagagat 1740

tcaaacctag acatacacat ccacaattgt cttaatctca gcagtactgg gaaagctttg 1800

tactcaactt aacctgtcat ttaacccttt ccactagttc tcccttaacc agactgcttc 1860

ctgtcttgaa acaaagaaaa aaccc 1885

<210> 23
<211> 494
<212> DNA

<213> Homo sapien

<400> 23
aagcgcgcgc attgtgatgg atctatattt taccctgtgc ttttctatag ctgtcctcaa 60

agcgtaaacc attccaaatt attttccaac gtagtgttat atgtgtgcag cagagctatt 120

tctgcctggg cattgccagt ccctgagcag gagggtctca cagtgaggtc tgcaggactg 180

taagtttggg gtctgactcc ctggccaccc tgtgtgggct gtgactgtct ctcagagcta 240

tacccgccct ttctctgctg gcagcccgac agagctggct caaccatcgg aggtcgcagg 300

ccaccagcca cgtggcacca ccatggcagc cttccaggtg aaggtgagac acacaaggca 360

tgacctgggg gccgaccgga tccccatcac aaacgccaca aacaccataa acacaaccca 420

ccctgatcag agactaagca gagaaagcag ggagaggacc tagagttact cagtaatgac 480

tcaggaagga gacc 494

<210> 24
<211> 1692
<212> DNA
<213> Homo sapien

<400> 24
gtcccccacc atggaagagg ccgggcccac ccactgcaag tcttctctga gccacgttct 60

caagtcttct ctgagccgcg ttctccaggt tgtgctgctg gagtcagttg gcatttcctc 120

caagcctgaa agtgtagtca gattcagaat gggcttttct agattcccct gtaagatctt 180

tcccctgctc ctggcaggag caccacacca tgggaacccc agggcccacg cagctgcccg 240

ggactggggg accaggacgt ggcacttctc acatgggtgg aaagatgggt ttacagaatg 300

gtggcatgga gacgctgtgg cctggcaagg atcaatgggg tggcatctgg cattagccat 360

caggaagact taaggctgaa gggacattgg gcagggagct ctcagggctg ctccacccgc 420

ccccagggtg acagcccata gtatcactta gggtgggact gagagtcacc tgggggagag 480

gagagaaggg gcccaacttc cccagcccct agtatcactt agggtgggac tgagagtcac 540

ctgggggaga ggagagaagg gacccaactt ccccagcccc tggcaccttc cctgcctttc 600

ccagtctttt accagagtca taagatggtc cttggctctg ggcaggcatg tggccctggg 660

gagctctggg gtcagaggtc aaggtgcttt gcatgtcagg caggcttgac ttttgcctgt 720

agaaagacta tagaaagatg gcaagctagg cctcttttct ggaaaagtgc caacagctga 780

taattttagg aaataatgtt ttgaatgtga agtgtgactt tttagaataa aaagacagga 840

agctcttaga aactgcaaga ttctaaatct aagcaaaagg ctatatttta ccctgtgctt 900

ttctatagct gtcctcaaag cgtaaaccat tccaaattat tttcaactag tgttatatgt 960

gttcagcaga gctatttctg cctgggcatt gccagtccct gagcaggagg gtctcacagt 1020

gaggtctgca ggactgtaag tttggggtct gactccctgg ccaccctgtg tgggctgtga 1080

ctgtctctca gagctatacc cgccctttct ctgctggcag cccgacagag ctggctcaac 1140

catcggaggt cgcaggccac cagcccgtgg cccacctggc agccttccag gtgaaggtga 1200

gacaaacaag gcatgacctg ggggccgccc ggctccccat cacaaacgcc acaaacacca 1260

caaacacaac ccaccctgat cagagactaa gcagagaaag cagggagagg acctagagtt 1320

actcagtaat gactcaggaa ggagacccta agcttctacc acatgccaga ctctgtgccc 1380

agtgcagcat aaacgtcctc agaaccagcc tggtcccagc ctggccgagc cggacgttcc 1440

tgggaaaggt tacaggagga gcagggccag gcccacagca cttttagaag cccatgaaaa 1500

tgtcttcatt tctcttcaaa tcacaaacaa aacgtgcaaa acccattctg gagtgcatct 1560

tttcactggc gaccaaccca gtcctaagat aaccttctta atagttctat ggaggaagct 1620

gcaaaggcag aagtgactac aacccacaaa agtcatgatg gagccctgac gtgtgtgtac 1680

acacacacta ca 1692

<210> 25
<211> 430
<212> DNA
<213> Homo sapien

<400> 25
acccagcgtc ccctggccag agccaccaga ggacagagct cccaatgagc ccagctgcta 60

gaaaagaagg tggagtccca ggcagaagag ttcttcaggc tgaacggaaa tgattccaga 120

gggaaatgca gatatgaaga aggagataaa gagctccaga aatggcaaat agcagggtga 180

gcctacgcga cttctctaac ggaagaaatt acctttaaaa cacacgtgca ggcttagagc 240

aaaagaaacc gtgccataag gtgtgagtaa gtgaagtgcc tgtgacacct acagatcaga 300

gaagcagagg cctccgggat ggcaaggcaa ggttgccgca tttcatatga agtgcacaat 360

catcataaaa gaatgcatta aatatacata tgtatgcatt caaattacac taacatcaca 420

tatatccatt 430

<210> 26
<211> 2603
<212> DNA
<213> Homo sapien

<400> 26
tgtttggtcc agtgaatctg cccaccaaca ccccgcctct caccatccac cagcccttgg 60

acccctagca ctgagctcac agtgaaaggg aatatttgct tgtaaataga aatagacgct 120
ggttagaaac caactggaaa gaatctttcg ttgaatagga gttaaaaaac aaggaaatta 180
acccactcct gggtatttct gaaactggca atctatgctt gtctaggacg gcccagacta 240
accctaatcg ccccgtcatc aacacagcag cacagcgttt cctccaggag aaacaccaag 300
atctcacgtc ccatccacag gctgaggctg ctgctcctgc aggaacctgg tgcagtgtag 360
caattccaca tcctgaaatt gctcatcaaa actcctatta aagtgtcaaa cagtgaatag 420
ctaaaatacc actttgcttg aacagtgaag aggttggaag gaaaacgtta actgtatcag 480
agaatatgga ctcctaacat acagggagtc aggttcattt tgaagtcact cttcttccaa 540
cagattcact aaggctcttt gtcaacacaa attgaaaacc gttaaaaaaa aaagtaatta 600
tgatgcttcc tgccctccat gaaaggacca catacagaca ccacgctcat atctgaggcc 660
ctggggtagc ctttaatggc ccagcagaat ggccagaacc gttagaggaa acatttaata 720
aagtctggag tcagagccct gcgggtctag ctggattcct ggaggtgcgg cccagaagcc 780
agcgggaggg aattggaggc cggaggctca aactgtcccc acttccacca agggcccctc 840
ctccaacagc ttccaggctg ccaaagcccc tgcatcacct ccagggtccc ctgggtccag 900
cctcatgctt cccataatga gtttttaaac cacaacgctg catcaggtga catctcttct 960
gcaggctgtg cgcgtctcca gggggaaggg ggctgtgtct ttgggacagt ttgtgctctc 1020
aatcacttga ctgctgacag gcacctcagc tgaatggtgt gatttatgca aagattgtgc 1080
tgaattattt aaagcattct ctatttaaag aacagagaat atttaattag cattctgctg 1140
tgcttaattg aagactcaca aatcaattaa aactgcttac cttttggcag ttcagtaact 1200
tcacagaaac ctcccaggaa atgcatccta ttcacagctg ggttcatcct atacccagcg 1260
acctgtggcc agtgtggcgc tgtgattaga ggcggctcag cgccttcaga ggagcggcct 1320
ggctgtgcgc acattagaga aaggcttcca tcgtcgttgg tcctctttct cacagggact 1380
ctggggtctt ggtgccggga gatgcaaccg cctctggcag cccggcttca ttttagggac 1440
agtgactatg gagaaaccca ggtctgaccc attttctcca gaggggaggg agccacaggg 1500
aaaggcccct tgttgctctg ttggccctga gtgcctccca ggaaaggtca gagcacagac 1560
tcagccctgg gagggccgag agatcccgct ggaccctgcc ctcctcgaca ctctggacaa 1620
gatgcagaga gtggggtcct ggcagcaaga tcccgtggga gtggggcctt ggagctcagg 1680
gccagaccga gggggtgctc attgctggct ctggcctaca gacacgttga cattggcacc 1740
acacgggcca actgaaaccc taagagaaaa cccagcgtcc cctggccaga gccaccagag 1800
gacagagctc ccaatgagcc cagctgctag aaaagaaggt ggagtcccag gcagaagagt 1860

tcttcaggct gaatggaaat gattccagag ggaaatgcag atatgaagaa ggagataaag 1920

agctccagaa atggcaaata gcagggtgag cctacgcgac ttctctaacg gaagaaatta 1980

cctttaaaac acacgtgcag gcttagagca aaagaaaccg tgccataagg tgtgagtaag 2040

tgaagtgcct gtgacaccca cagatcagag aagcagaggc ctccgggatg gcaaggcaag 2100

gttgccgcat ttcatatgaa gtgcacaatc atcataaaag aatgcattaa atatacatat 2160

gtatgcattc aaattacact aacatcacat atatccatta gactttatca aaattaaaat 2220

cttctgttca tccacataaa acgatgtcac ttactgcaaa aaatattctc aaatatttat 2280

ccaagtgctg agatccagaa taagtaaccc ctaaaatttc ataataaaac aacttggtga 2340

aacaacggtc aaaggatttg aacacttcgc caaatgatgg caaataaaca caagaaaaag 2400

tgctcgacag actcgagcac caggaagatg cgtcgtaaac accaacaaaa accaccacac 2460

acacccacag tagccaaaat ctataaaact ggtggcacca aacgtgaggg aggatgtggc 2520

ccacccagca ctgttgctgt gcattcttgg tgagaacacc taagacgtcc cctcaatggg 2580

attagaaaac cacaaggcag gca 2603

<210> 27
<211> 614
<212> DNA
<213> Homo sapien

<400> 27
acatatattt aaagggaaga tggatacaat ttgtttttat tatataaatc taggtaaggt 60

gaaatgcttt tgtcaacaaa aatacagtgt agtgaatttt atatttgtcg cttgattagg 120

taaactgaaa actaacaata gaaatattat tttactgcat tgaaatacca tgaactttca 180

gacttgttag ttctacaagc agttgtgcta ccttaatttt gtgtttccag aaataaaaat 240

taaccttagt tatgctgtca tttttaacta ataaaaaaag tataattcat aaaacttttg 300

gctttataag ataattataa aattatatat ttttttctgt ttttgtgggg ttgggaaaac 360

attttcttat ttctattcac tcttcaaatg caggtctcat aatatgtgtc aatgatataa 420

gatgatggaa gacttctgta ataaaaacat atgtcattat cttcaatttg ttcaataaat 480

aatttaactg tgaaacaaca aaaaaaaaaa ccaaaaaaaa aaaaaaaaaa acaaaaaacg 540

ggggggggcc accggggcaa agggggcccc gggggaaggg ttcccgggca aatccccata 600

agagcaaaaa acat 614

<210> 28
<211> 1134

<212> DNA
<213> Homo sapien

<400> 28
gcacgaggat tggtcaaagt agtattctct tgaagttcta gtcaatttaa tttgatccaa 60

taagtttttc tgaatctcct ttttaagttc caagaaattc tattataaat aagtgtactt 120

ttaccaattc cattgtataa gcaaacagac accttttaga aaaggataag taatcatcaa 180

tttgtttttt ttaaaaaaaa acaatttcca gactactaaa tttggcataa gaataattct 240

tttaaaatgc aacatacttt aattagtttt tttggtatat gcataagatg tgaactttcc 300

tattgatatc actttatatt aatagagatg tacatttctt tctatgccgt ggctagagca 360

aaagttaata atgattattt acacaattga tttaatttct taggatatgt ataatattgg 420

atattatatc tgatttaaaa atactattcc atacattttt tttttcagga gataaaacat 480

agggaaaggt tttcatgtga attctttgta tcactttgaa gtacatatat ttaaagggaa 540

gatggataca atttgttttt attatataaa tctaggtaag gtgaaatgct tttgtcaaca 600

aaaatacagt gtagtgaatt ttatatttgt cacttgatta ggtaaactga aaactaacaa 660

tagaaatatt attttactgc attgaaatac catgaacttt cagacttgtt agttctacaa 720

gcagttgtgc taccttaatt ttgtgtttcc agaaataaaa attaacctta gttatgctgt 780

catttttaac taataaaaaa agtataattc ataaaacttt tggctttata agataattat 840

aaaattatat atttttttct gtttttgtgg ggttgggaaa acattttctt atttctattc 900

actcttcaaa tgcaggtctc ataatatgtg tcaatgatat aagatgatgg aagactttgt 960

aataaaaaca tatgtcatta tcttcaattt gttcaataaa taatttaatg tgaaaaaaaa 1020

aaaaaaaaaa ccaaaaaaaa aaaaaaaaaa acaaaaaacg gggggggcc accggggcaa 1080

agggggcccc gggggaaggg ttcccgggca aatccccata agagcaaaaa acat 1134

<210> 29
<211> 1139
<212> DNA
<213> Homo sapien

<400> 29
cgaggtaccc attataatta ctaaactgtg aagtcactat tattagtatc tgaccagcta 60

tacaaaacat catcaatttt acttttgaca caaaaggtag taaaaatcgc aaacgataaa 120

gaagacacta ctcattaaaa gtcatgttta ctaatccagc accataattc cagtctcaga 180

acctcccatg cagattggaa agggattatg ggaacgaggt gagtatgtag gacatgtcgg 240

cgctagtaac atcaaattga cggccccata tttgctcgct tcacaagaca aaaaacacag 300

ggtcctccca aagtaagcag aagatgacat gacggcatgg agacgaaaaa caaaacgcta 360

gcgcgctaaa tcaatggtca atagctgcaa aaccatctga tgacaactag ggtaacttcc 420

cgtgtcaacc aaaaattcac aaacaagtaa gcactacctg tagaacagac acgaagtcac 480

gcaaacctac actttgagca cgcctgacca gagatccgag cacactcccc gacccaccaa 540

cacacagcag gccacgcggt agagagaaca agaatacaaa ggacaagcga gtagctgtag 600

aagcgatgag agagagcgta cgtagagatg ggggaggaac accacgtagg agcagaactg 660

ctgcactgcg tgcacacgcg acgcgaacag acgaaactac acgaagacaa aaggaaaagg 720

aaaggatggg accagagggg agagccaagc atgagagaca caccaaaagg cacccgcacg 780

ctgcatggcg aagcgagaag aacagcagat aaccacaaaa aaaagcacac acggtgggac 840

atacacacca gagggggagc atcagacaca gggacaaacc actaaagcag gagaacatgg 900

cgcgaaagga ctgaactaaa cagcacaaac acgcaacgag cagcgaacag ccgatcatag 960

gcgtgacacc cgactacagc aaaagaaacg gagaagttat cgacacaagg gatgacaagg 1020

aaacaggcta atggcccaag gagaggaaca ataagatgga tgagcacagt agggcgaaca 1080

agggataacc caagtgaaga aacagtgaag aagaggaatg cacacaataa gaacgcaaa 1139

<210> 30
<211> 235
<212> DNA
<213> Homo sapien

<400> 30
agtgtttgca acagcaccat ttgtcaaatt caaagatgct caaaaggtgt tccctacttt 60

gcatgagagg gagagctttg taacaggaaa ttgtataagg caaactctct attcattcct 120

aaggcctctg ttcattccta atgtttacat ggttctctac tctgaagggc accaacatgg 180

acctcacctt cttaacatgg aaaatcaaaa tctaaatgaa ttaccattaa aagga 235

<210> 31
<211> 2171
<212> DNA
<213> Homo sapien

<400> 31
ctgcattttt ctgtcattct cttcatttgt tttaaggtgg aaaattttct tacagttgat 60

gcaaagtatc aackacttta ccctaccttc tccccttttta gatgggttct tcctgagttt 120

tggagtcttg tatgattatc agtattcccc tgtcaaaatc aaatctattc aggtttcttc 180

actgttgaga acacctaaat gtttttattt ttgagaagtg gggacagagt ctcactatgt 240

cacccaggct ggagtgcaat ggcatgatct cagctcactg caaccttcgc ctcctgggkt 300

caagcgattc tcctgcctcc gcctcctgag tagctgggat tataggcacg caccaccacg 360

cccagytwwt tttttgtatt tagtagagac agagtttcac catgttggcc aggctggtct 420

tgaactcctg accttgtgat ccacccacct cggcctcccg agggtgctgg gattacaggc 480

atgagccacc acgcttggct aagaacacct aaatttttat gtttcttggc tcaaaaacca 540

gttccatttc taatgttgtc ctcacaagaa ggctaattgg tggtgagaca gcaggggagg 600

aggaagagct gtggtttgta acttgttcaa ctcaggcaat aagcgatttt agctttattt 660

aaagtcttct gtccagcttt aagcactttg taagacatgg ctgaaagtag cttttctatc 720

agaattgcag atagtcatgt tgggctaaca gtcaattgga tatattcctt tacctcacat 780

gaccccagca actgtggtgg tatctagagg tgaaacaggc aagtgaaatg gacacctctg 840

ctgtgaatgt tttagagaag gaaattcaaa aaatgttgta actgaaagca ctgttgaata 900

tgggtatcgg ctttcttttt cactttgact cttaacatta tcagtcaact tccacattaa 960

tgaaagttga ccatagttat ttccaaataa aaagaaacca actcttacca ggtcttggac 1020

tgtgatgtca tattattcag ttttatgctt gttcctgagc agaactcata agagtgacat 1080

agtcagctgc tgacggcacc tcagccacgc cactcttact cagttcagtg ggtgtgcttg 1140

cgtggtagga tgtggtgcag ccctctctac gctcttctat ttttggtata tttcctatct 1200

aaccttcaaa tagcttccaa ttcttttttt cttggactgg cttcattctg aatttgtgct 1260

aaaataatct ttcataaaga gacctcagtt tatagcgtaa cagactacac aatgcactga 1320

tgttttcata atgtttaagg gacccactgc aagaagcttg ctgcctcctt ttaattgtat 1380

tcatttagat tttgattttc catgttaaga aggtgaggtc catgttggtg cccttcagag 1440

tagagaacca tgtaaacatt aggaatgaac agaggcctta ggaatgaata gagagtttgc 1500

cttatacaat ttcctgttac aaagctctcc ctctcatgca aagtagggaa cacctttga 1560

gcatctttga atttgacaaa tggtgctgtt gcaaacactt ttttttgag atgaagtctc 1620

gcggttgtca cccgggctgg agtgcagtgg cgtgatctcg gctcactgca acttccacct 1680

cctgggttcc agccagttct cctgcctcag cctcccaagt agctgagatt acaggcgcct 1740

gccaccccac ctggctgatt tttgtaattt tagtagagac ggggtttcac catgttggcc 1800

aggctgatta actcctgacc tcaggtgatc cacctttctc ggctcccaaa gtgcttggga 1860

ttacgggtgt gagccaccgt gcccggcctg caaacacatt ttaattgaca acactagggc 1920

tgttgtacaa aatagtaatg atagccatgg aagttttacc ttattctgtg agaagtgttc 1980

ttaaacttat taaagtgtct aaaactaagg ttagtgcttc taaaggaagt ggccggttct 2040

```
cctaagaagc aattatcact gtccctgact ttgtctggtt ggtttggttc cccctgtccc   2100

cgattggctc tggtgtcctg ctttgccgcg gttcttttaa gccagcgcgg gttatttttt   2160

gaaaacctcg g                                                        2171


<210>  32
<211>  192
<212>  DNA
<213>  Homo sapien

<400>  32
gcgtggcgcg gccgaggtac tgtctctaca gccattgaga agccattcag tgccctggta     60

gggacctgag actttccaga attcacacag cagtctatga tccctcaaat gtaagaggac    120

agggggtcag cctatcttca cctctcagtg aatgtggagg gccaagcaat atgacttgca    180

aacctaagct ag                                                        192


<210>  33
<211>  2641
<212>  DNA
<213>  Homo sapien

<400>  33
tttttttttt ttcttttcca agttatttaa tttacagcat cagtctccaa atataataat     60

attaagatag cagtttagaa attaactttt tttcagatca ctctaacata aaatctctca    120

actgaatctc tagtttgtct cattttgtta agagctttaa tattacatgg gaagttcaga    180

gacttctatt tccatccctc aacatgtagt gacagtcaac atgtcaggct ctgtagcacc    240

gtgatatccc agcaccagac cactccagcc accctctcat tcaaagaagg gctacaagat    300

atggctggac tactcgaatc acatctgatc ttaatcaatc caggtataga aagttgtact    360

ataaagaata ctttccaaaa ttgttcactc aaataaaaac agatcaagtc attacagagc    420

atttttccat tttaataaga ataacagacc tactcaaggt aattttattc tgtttattta    480

aataaggata agactactta aaagactttt tacatacaaa aatgtacaag gttaaacttt    540

tctgtactga attacaaaac ctgcacaagc atgtaataaa agagcacact taaaaacatt    600

ctgaccatta tttagcctct aaaaattact gaagttcaac agtagtaaat agaggaagct    660

cttacatata tatatatata tatatatata tatatgattt aatctactgg cagttttact    720

taatgtaagt atttaaaagg tcacattgct attgaatgag tctctagatc aattttagaa    780

ttgtctctca aaacttaagt caaccaaaat attatttcaa atagtaattc caattctgaa    840

gaattttaat accagcaaat atattatggc ctcatagtag taactgaacc aactttccaa    900

agtgcctggt agctgtccag atgaattagg ctgctttgga aaactgtact gtctctacag    960
```

ccattgagaa gccattcagt gccctggtag ggacctgaga ctttccagaa ttcacacagc 1020
agtctatgat ccctcaaatg taagaggaca gggggtcagc ctatcttcac ctctcagtga 1080
atgtggaggg ccaagcaata tgacttgcaa acctaagcta gaagcttggg atctacagta 1140
aggaggaagg agaattaaag tagagaaaga aaatgtataa ggagaaaggg aaaagaagga 1200
acaaagaggg aaaagaagaa aaaacaagga tgcctgctaa tggcaggaag tggtaaagtg 1260
cctataacta caacttacaa gccacccact aattctaatg ccattcattt gcctactcca 1320
ataataagaa aagctggctt tactggaata tagaatctag agcaacatta cccgcctcat 1380
gttagtgagt aactagtatt ctaaagttgt ttgccataca tatcaagttc ttctaacctt 1440
tgaagcaaac caaaacactt caaaactcag ggctcccagg gctgctgctc cagattccca 1500
gcattcagca tgcttcatta tgtggagaaa gacatttcaa gacaagctgt atctatacac 1560
cttcagaagg aacaaagctc taagaaggtg ggattatgtt aacacatagt acatggttta 1620
gcgtttctcc acatttcaaa ctcaaaatag ctcaataata tgctgctaca tgagcattga 1680
ttctgaatgt tcataatata aacttcaatt tgaagcaaca atgttacaca gttcagctgt 1740
tattaccaac ctactctgta agttaaaata caaataaaat attaatttta ttgagtaact 1800
aaaaataagt tcccactgac ttaaaatcgt caaatggcta actctctctc aactaagaga 1860
gcaacacaga tggaagcaga gaggacaact gaatataaaa taaaatttgt caatctactc 1920
tataatctgc acttttaaaa tccccttttg catatatgta tgtataggat cacagttgcc 1980
caccaacatt atgtctgtca gccctgcaga taacaattta ctgtaacgtt aacaatttat 2040
gcaatactta gtatgtttta tcttatgtgt acagatttac agtttggaat aaaggcagaa 2100
tgattaaaaa ctattgggtt aaagtcttag tatggtactt acctgcaagg ctgaattaat 2160
tttttggaag gctattcaat agctgaacta aaatgcttgt ttaacaaatc aaaagaggaa 2220
taagactact ttaaaacata ttgaaaaagg taaatcccaa tttgaagatc aatcatataa 2280
cgaaaaaagt atgaagtatc ctttgctctt gcttagaaac acatagcaga acagtagaaa 2340
ctagaactca tgaatataag gtaaacccta ttttcccact gatttccatt atacaattgg 2400
agtgaaaata ccactcaaac aaaaataaac aaaaaatctt agcaggtaat tctgtgtaga 2460
acagccatgt gggaattgtc tatattacag ctgcagggaa tctcatgtaa gctaggagtc 2520
catcttccta tgttgcactc tgcagtgact tctgactccc agtagctcct ctattgccta 2580
ctccatatta cgctaatttt tgccccctga ctgctatgct tcctgggact cttattaaat 2640
t 2641

<210> 34
<211> 434
<212> DNA
<213> Homo sapien

<400> 34
atttccttat acacacaccg aatcagaata tactttcagt tctacaattt gacaatacac 60

atagctgatt tatagcaagt gtgccatgaa ctgagggttt gtttagtttg tttttgcagg 120

gctgccaata tgctgtcttc acgggacggt aaagaaagta tcacttgggc cgcatctaat 180

atgaaatact gaaggtgggt gtagagaggg tgctagggct ttgaacagcg gcacttcctt 240

tctgagagag agaaaacatc atgctccccc cgcgccgaac tcattttaca ggttgattgg 300

gtgaacaatt cttggcaggc cctgagctag tctgggtatc ctgagtcaag agagaggccc 360

tgcctctgag gtaaagtgtc tctcatctgc ctaagtttgc ttagaaactt tggcttatga 420

aagattaacc taag 434

<210> 35
<211> 197
<212> DNA
<213> Homo sapien

<400> 35
tctgagacaa tagggcatgg gtcctctaat tcatctcgag cggcgcatgt gatggatagc 60

ggcgcccggg cagggaaacc cctactggac cctgtgtgtc tgccagcctg gagcctttgt 120

ctccagccct gcctttattc ctccttgcct ccacaccagc ctccccttgc ttctccttac 180

agactatcca agaagtg 197

<210> 36
<211> 3414
<212> DNA
<213> Homo sapien

<400> 36
atgggggatt tcgcagcccc cgctgctgcc gcgaatggca gtagtatttg catcaacagt 60

agcctgaaca gcagcctcgg cggggccggg atcggtgtga ataatactcc caatagtact 120

cccgctgctc cgagtagcaa tcacccggca gccggtggat gcggcggctc cgggggcccc 180

ggcggcggtt cggcggccgt tcccaagcac agcaccgtgg tggagcggct ccgccagcgc 240

atcgagggct gccgtcggca ccacgtcaac tgcgagaaca ggtaccagca ggctcaggtg 300

gagcagctgg agctggagcg ccgggacacc gtgagcctct accagcggac cctggagcag 360

agggccaaga aatcgggcgc cggcaccggc aaacagcagc acccgagcaa accccagcaa 420

gatgcggagg ctgcctcggc ggagcagagg aaccacacgc tgatcatgct acaagagact 480

gtgaaaagga agttggaagg agctcgatca ccacttaatg gagaccagca gaatggtgct 540

tgtgatggga atttttctcc gactagcaaa cgaattcgaa aggacatttc tgcggggatg 600

gaagccatca acaatttgcc cagtaacatg ccactgcctt cagcttctcc tcttcaccaa 660

cttgacctga aaccttcttt gcccttgcag aacagtggaa ctcacactcc tgggcttcta 720

gaagatctaa gtaagaatgg taggctccct gagattaaac ttcctgtcaa cggttgcagt 780

gacctggagg atagcttcac catcttgcag agcaaagacc tcaaacaaga acctctcgat 840

gaccctactt gcatagacac atcagaaaca tctctttcaa atcagaacaa gctgttctca 900

gacattaatc tgaatgatca ggagtggcaa gaattaatag atgaattggc caacacggtt 960

cctgaggatg acatacagga cctgttcaac gaagactttg aagagaagaa ggagccagaa 1020

ttctcgcagc cagcaactga gacccctctc tcccaggaga gtgcgagcgt gaagagcgac 1080

ccctctcact ctcccttcgc acatgtctcc atgggatctc cccaggcgag gccttcttct 1140

tctggtcctc ccttttctac tgtctccacg gccactagtt taccttctgt tgccagcact 1200

cccgcagctc caaaccctgc aagctcacca gcaaactgtg ctgtccagtc ccctcaaact 1260

ccaaaccaag cccacactcc aggccaagct ccacctcggc ctggaaatgg ttatctcctg 1320

aatccggcag cagtgacagt ggccggttca gcgtcagggc ctgtggctgt gcccagctct 1380

gacatgtctc cagcagaaca gctcaaacag atggctgcac agcagcaaca aagggccaaa 1440

ctcatgcagc agaagcagca acagcaacag cagcagcagc agcagcagca gcagcagcag 1500

cagcaacagc agcagcagca gcagcaacag cactcaaatc agacttcaaa ttggtctccc 1560

ttaggacctc cctctagtcc atatggagca gcttttactg cagaaaaacc aaatagccca 1620

atgatgtacc cccaagcctt taacaaccaa aaccctatag tgcctccaat ggcaaacaac 1680

ctgcagaaga caacaatgaa taactacctc cctcagaatc acatgaatat gatcaatcag 1740

cagccaaata acttgggtac aaactcctta aacaaacagc acaatattct gacttatggc 1800

aacactaaac ccctgaccca cttcaatgca gacctgagtc agaggatgac accaccagtg 1860

gccaacccca acaaaaaccc cttgatgccg tatatccagc agcagcaaca gcagcagcaa 1920

cagcaacagc agcagcagca gcagcagcag ccgccacctc cacagctcca ggcccccagg 1980

gcacacctga gcgaagacca gaaacgcctg cttctcatga agcagaaagg agtgatgaat 2040

cagcccatgg cttacgctgc acttccatcc cacggtcagg agcagcatcc agttggactt 2100

ccccgaacca caggccccat gcagtcctcc gtgcccccag gctcaggtgg catggtctca 2160

ggagccagtc ccgcaggccc cggcttcctg ggcagccagc cccaagcagc catcatgaag 2220

cagatgctca ttgatcagcg ggcccagttg atagagcagc agaagcaaca gttcctgcgg 2280

gagcaaaggc agcagcagca gcagcagcag cagcagattt tggcggaaca gcagttgcag 2340

caatcacatc taccccggca gcacctccag ccacagcgga atccataccc agtgcagcag 2400

gtcaatcagt ttcaaggttc tccccaggat atagcagccg taagaagcca agcagccctc 2460

cagagcatgc gaacgtcacg gctgatggca cagaacgcag gcatgatggg aataggaccc 2520

tcccagaacc ctgggacgat ggccaccgca gctgcgcagt cggagatggg actggcccct 2580

tatagcacca cgcctaccag ccaaccagga atgtacaata tgagcacagg catgacccaa 2640

atgttgcagc atccaaacca aagtggcatg agcatcacac ataaccaagc ccagggaccg 2700

aggcaacctg cctctgggca gggggttgga atggtgagtg gctttggtca gagcatgctg 2760

gtgaactcag ccattaccca gcaacatcca cagatgaaag ggccagtagg ccaggccttg 2820

cctaggcccc aagcccctcc aaggctgcag agccttatgg gaacagtcca gcaaggagca 2880

caaagctggc aacagaggag cttgcagggc atgcctggga ggactagtgg agaattggga 2940

ccattcaaca atggcgccag ctaccctctt caagctgggc agccgagact gaccaagcag 3000

cacttcccac agggactgag ccagtcagtc gtggatgcta acacgggcac agtgaggacc 3060

ctcaacccag ctgccatggg tcggcagatg atgccatcgc tcccggggca gcaaggcacc 3120

agccaggcga ggccaatggt catgtctggc ctgagccagg gagtcccagg catgccagcg 3180

ttcagccagc ccccagcaca gcagcagata cccagtggca gctttgctcc aagcagccag 3240

agccaagcct atgagcggaa tgcccctcag gacgtgtcat acaattacag tggcgacgga 3300

gctgggggtt ccttccctgg cctcccggac ggtgcagacc ttgtggactc catcatcaaa 3360

ggcgggccag gggacgagtg gatgcaggag cttgatgaat tgtttggtaa cccc 3414

<210> 37
<211> 678
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (310)..(611)
<223> a, c, g, or t

<400> 37
tcataatgct gtcgagcggc ccgcagtgtt gatggatcgg ccgccgggca ggtacctgct 60

gtgtggcagg ctctgggctg gggctttat tcagcttcct cagcctgctt cgacttcccg 120

attagagagc taatgtgaat caccaaccct gtgatgcctc ttgagatgag agttcagatt 180

```
tcccaagaag atctaagcag ttggtccaaa ttgtagttca ctagcaaatg acccagtgct     240

gtccctgtgg tgtgtttatg acatgatgga agatgctgcc ttcaaaagtg tccacttgta     300

agaagatgtn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     360

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     420

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     480

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     540

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn     600

nnnnnnnnnn nacacacaaa aaaaaggtgg gggaaccagg gcaagcggtc ccgggggaaa     660

tggtttccgc acattcca                                                   678


<210>  38
<211>  461
<212>  DNA
<213>  Homo sapien

<400>  38
gtgcggccga ggtaccaact gacatttcag tttttctgtt tgaagtccaa tgtattagtg      60

actctgtggc tgctctcttc acctgcccct tgtggcctgt ctacaattct aaatggattt     120

tgaactcaat gtcgtcgctt ctggtttcct gcatatacca atagcattac ctatgacttt     180

tttttcctg agctattttc actgagctga gctaatgaac taaaactgag ttatgtttaa      240

tatttgtatc aaatacataa aaggaatact gctttttcct tttgtggctc aaaggtagct     300

gcattttaaa atatttgtga aaataaaaac ttttgttatt agaaaaaaaa aaaaaaaaaa     360

aaaaaaaaaa ggcttggggg aaacccgggg ccaaaagcgg tgtcccgggg gggaattggt     420

ttctccggtc caaattcccc aaaaaaatcg agaagaaaag t                         461


<210>  39
<211>  633
<212>  DNA
<213>  Homo sapien

<400>  39
caacaccatc tttttttttt tttttttttt ttgagacaga ttcttactct gcactccagc      60

ctggtgacag agcgagattc catctcaaaa aaaaaaaaac agtatgcacg tacaaatttc     120

ttaacctgtt atcaatgtct gagctacata attatctttc tagttggagt ttgttttagg     180

tgtgtaccaa ctgacatttc agtttttctg tttgaagtcc aatgtattag tgactctgtg     240

gctgctctct tcacctgccc cttgtggcct gtctacaatt ctaaatggat tttgaactca     300
```

atgtcgtcgc ttctggtttc ctgcatatac caatagcatt acctatgact tttttttttcc 360

tgagctattt tcactgagct gagctaatga actaaaactg agttatgttt aatatttgta 420

tcaaatacat aaaaggaata ctgctttttc cttttgtggc tcaaaggtag ctgcatttta 480

aaatatttgt gaaaataaaa acttttgtta ttagaaaaaa aaaaaaaaaa aaaaaaaaaa 540

aaggcttggg ggaaacccgg ggccaaaagc ggtgtcccgg gggggaattg gtttctccgg 600

tccaaattcc ccaaaaaaat cgagaagaaa agt 633

<210> 40
<211> 536
<212> DNA
<213> Homo sapien

<400> 40
ggggccgccc gggcaggtac ttgacagtgt tatctgtcac ttatttaaaa aaaaaacaca 60

aaaggaatgc tccacatttg acgtgtagtg ctataaaaca cagaatattt cattgtcttc 120

attaggtgaa atcgcaaaaa atatttcttt agaaacataa gcagaatctt aaagtatatt 180

ttcatataac ataatttgat attctgtatt actttcactg ttaaattctc agagtattat 240

ttggaacggc atgaaaaatt aaaatttcga tcatgtttta gagacagtgg agtgtaaatc 300

tgtggctaat tctgttggtc gtttgtatta taaatgtaaa atagtattcc agctattgtg 360

caatatgtaa atagtgtaaa taaacacaag taataaatga agtgtttgtt ataaaaaaaa 420

aaaaaaaaaa aaaaaaaaaa aaaaaaaagg gtggggggaa cccggggcca aaaggggttc 480

cgggggggaa attggtttcc gggccaaaat ttccaacaat ttgggagaaa aaaggt 536

<210> 41
<211> 1206
<212> DNA
<213> Homo sapien

<400> 41
gtactctccc aaatgcagcc taatcttagt aaccttgaag tttatcattc tttaaaacta 60

aatagaatac caatggttta gatattccaa caaagaatgc tagaaacaaa tgtctaatct 120

cgattattag ctttaccaac cctgtgaaca ctgaggttgc agaactgcca ggttaatccc 180

tgtggcctag actactgagg attctgatag cacatgtaag actaagcact cttcaagctg 240

taataaagca tccacatgta tctgtgatga ttttcattgc tttagcattg cagccatgta 300

acaactgcag aaagaaggta tttttaaaaa tacaatagac tacacttttt ggatcacaga 360

gaaatacaga tgcactctga gactgcctat gtttataaac atgttgtgtc ccctaactga 420

agtgacaggt cttctggaat tgacattaag aagtgtggat agtcatatca cacgcaatgt 480

atttgttttc agcagtgagc agaccgtaca ggagcagcac accaggagcc atgagaagtg 540

ccttggaaac caacagggaa acagaactat ctttatacac atcccctcat ggacaagaga 600

tttatttttg cagacagact cttccataag tcctttgagt tttgtatgtt gttgacagtt 660

tgcagatata tattcgataa atcagtgtac ttgacagtgt tatctgtcac ttatttaaaa 720

aaaaaacaca aaaggaatgc tccacatttg acgtgtagtg ctataaaaca cagaatattt 780

cattgtcttc attaggtgaa atcgcaaaaa atatttcttt agaaacataa gcagaatctt 840

aaagtatatt ttcatataac ataatttgat attctgtatt actttcactg ttaaattctc 900

agagtattat ttggaacggc atgaaaaatt aaaatttcga tcatgtttta gagacagtgg 960

agtgtaaatc tgtggctaat tctgttggtc gtttgtatta taaatgtaaa atagtattcc 1020

agctattgtg caatatgtaa atagtgtaaa taaacacaag taataaatga agtgtttgtt 1080

ataaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaagg gtggggggaa cccggggcca 1140

aaaggggttc cggggggggaa attggtttcc gggccaaaat ttccaacaat ttgggagaaa 1200

aaaggt 1206

<210> 42
<211> 209
<212> DNA
<213> Homo sapien

<400> 42
ccgggcaggt ggaacttagt gggcagcatt acgggcagcg ctaaggaacc atttaaagta 60

agacaagtcc acacagctgt ggtgcttttc tacgagtctt gttccaactg ctgcataaca 120

atagaatgtt ggaagcagga attagtttta aagtaagact tcagaagtgg aaacaaattt 180

gatatttatt tttataatga tataatagc 209

<210> 43
<211> 706
<212> DNA
<213> Homo sapien

<400> 43
gaaccctcca aaacatctga aaagcaaatt tggggggatg aggaagtgag atgatgactt 60

gattctcctt ctaggaagaa tagaggaacc cttctggcaa aatttcaagc atctacaaga 120

ggaggttttc cagaaaataa agacactggc tcagctctca aaggatgttc aggatgtcat 180

gttctacagt atcctggcca tgctcagaga cagaggggct ctacaggacc tgatgaacat 240

gctggaattg gacagctcag gtcatttgga tggccctggt ggtgccatcc taaagaaact 300

tcaacaggat tcaaaccatg catggtttaa cccaaaggac cccattcttt atctccttga 360

agccataatg gtgctgagtg acttccaaca cgatttgctg gcctgttcca tggagaagag 420

gatcctgctt cagcaacagg agctggtaag gagcatcctg gagccaaact tcagataccc 480

ctggagcatt cccttcaccc tcaaacctga gctcctcgcc ccactccaga gtgagggttt 540

ggcatcacct atggctgctg gaggagtgtg gccttaggac ggagctggat aaccccaggt 600

caacctggga tgtagaagca aagatgccct gtctgtcctc tatgggactc tctcgttgct 660

gagcagtggg tgaaggctaa gcctccctga tgggagcagt cagaaa 706

<210> 44
<211> 1298
<212> DNA
<213> Homo sapien

<400> 44
atatgaagtt aaaaccagag ctatttctga cacagcaatt tttgagcggg catttgccaa 60

aatacgaaca agttcacatc ctcccagtag gtgagtgtga gtttgctgga ggtgggggtg 120

gggatcccat cctgcacaca tggggtaagt agggcagatt gcccctgcct cgcctttgcc 180

accaccgccc tagggcctgg cgtttggtca tgtggaatgg gaagggtcca gaaagctgag 240

aacatggagg atgaatggga atgggggcag gaagaagttg agtaagaggg aggaggtggt 300

aggagagcag aaccctccaa aacatctgaa aagcaaattt ggggggatga ggaagtgaga 360

tgatgacttg attctccttc taggaagaat agaggaaccc ttctggcaaa atttcaagca 420

tctacaagag gaggttttcc agaaaataaa gacactggct cagctctcaa aggatgttca 480

ggatgtcatg ttctacagta tcctggccat gctcagagac agaggggctc tacaggacct 540

gatgaacatg ctggaattgg acagctcagg tcatttggat ggccctggtg gtgccatcct 600

aaagaaactt caacaggatt caaaccatgc atggtttaac ccaaaggacc ccattcttta 660

tctccttgaa gccataatgg tgctgagtga cttccaacac gatttgctgg cctgttccat 720

ggagaagagg atcctgcttc agcaacagga gctggtaagg agcatcctgg agccaaactt 780

cagatacccc tggagcattc ccttcaccct caaacctgag ctcctcgccc cactccagag 840

tgagggtttg gccatcacct atggcctgct ggaggagtgt ggccttagga cggagctgga 900

taaccccagg tcaacctggg atgtagaagc aaagatgccc ctgtctgccc tctatgggac 960

tctctcattg ctgcagcagc tggctgaggc ctaagccctc cctgatgggc agtcagtcca 1020

gagatgctgg ccctcgccca gtctatgctg tgagtgtcct tatgggtgca agagataggg 1080

ctgtgcctct ctgcgtttcc aggtggagta gagacagtaa tgggtagaga ctttaggaaa 1140

tgttttgggg tggtggaata ctctatatat tgacaagagt ttatatattg acaagagttt 1200

atatatttgt caaaactcct caaatagtat gttaaagacg taagcgtttc actatgtata 1260

aattttactt caaaataata aaaacaaata ctgactct 1298

<210> 45
<211> 531
<212> DNA
<213> Homo sapien

<400> 45
acaacattca aacaaccagt ggtgaggttg taaatcaaat gagagaggag gaactgatcc 60

gggtagcagg aacacatttc caagtaaaat ttgcaacaga gcatgttgag atcatggttt 120

taatttatga atggcattat tatctttaaa ctattatttt ccaagctcat atatggcctt 180

tttgaaggtt ttccgaatgt tacatttgat tttaagatct aatccaaaat gaaatataga 240

atgtgcttag ttttctataa aaatgccaat gactatctct taaattagtc aaggaaagac 300

aaattaccaa aattcaaact tatttgaatt atttttaagt gattccaggc aataaataca 360

tagaacccat ggaaagtttt agcttcaaat cacaaaattg caaaaaaaaa aaatggtaaa 420

tggctaaaca taagggggt tatggaaaat attgggtcac cttaattata ggtttaaatg 480

ccacaaacaa tataataata gttttaactt actttttcg attactaagc a 531

<210> 46
<211> 469
<212> DNA
<213> Homo sapien

<400> 46
taacgccatc agctcgctgc ttaaagccgt gtttgcgtct cattttctca aagaaatctg 60

ctttagtttg agattacagt ttatcaaatg ttaaggcttt gaccccaaaa tctggtccca 120

gaaagacagg aaggccagct aagaggaggt tttcagagtg cgtagaaagg ctgctctgtg 180

cttcggcatt tgttctggaa gtgcttcttc ggttggcaaa gattcctagc aaaacctttg 240

actggaggct ttacagggcc atacacccaa tatcactaat gacagtgttg taaaatagct 300

tttgtgcacc atgcttagga ttcaaggagg ataaagtata tctttctaaa gttatacttt 360

agaaactgtc attccatgtt gaaatgataa acattccatg tttatctttt gtgtaagaag 420

taaaaaagca aaaattcatt gcatcaaagt aggtcaggca ctgctaaag 469

<210> 47
<211> 483
<212> DNA
<213> Homo sapien

<400> 47
aaaccgagtt ctggagaacg ccatcagctc gctgcttaaa gccgtgtttg ctctcatttt 60

ctcaaagaaa tctgttttag tttgagatta cagtttatca aatgttaagg ctttgacccc 120

aaaatctggt cccagaaaga caggaaggcc agctaagagg aggttttcag agtgcataga 180

aaggctgctc tgtgcttcgg catttgttct ggaagtgctt cttcggttgg caaagattcc 240

tagcaaaacc tttgactgga ggctttacag ggccatacac ccaatatcac taatgacagt 300

gttgtaaaat agcttttgtg caccatgctt aggattcaag gaggataaag tatatctttc 360

taaagttata ctttagaaac tgtcattcca tgttgaaatg ataaacattc catgtttatc 420

ttttgtgtaa gaagtaaaaa agcaaaaatt cattgcatca aagtaggtca ggcactgcta 480

aag 483

<210> 48
<211> 600
<212> DNA
<213> Homo sapien

<400> 48
tccatttctc atggcttgct catcttccgg cttcaggctc tgacttcatc tcaggatggg 60

atcggtgtgt gtctgttttc atagatccac tacatcagaa gtatctttac atctctgtat 120

ctttacatcc caaggtcaag gccctggcaa cctcagaggt tcccatagct tcagtcttcc 180

ccaaaccatg ccacttcctc ccatttcttt gggtcaggaa tctggctttt gttttccata 240

tttctttttc ccaagacatt gggaggcatc tggtgaacaa caccaataaa acagttctct 300

ccccaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaac aaaaaacgaa 360

gaacaaagaa cagagaaaaa aaaaaacaag aaacaccaaa aaacaaaaaa gaaaacgcgg 420

ccgccagcgc acgcgcgagg gcgccggagc acaccctgtg gccagcccgc gagcgagaag 480

ggagcgggcg gggcgggcgg gaccggagac ccaaggaggg cgcagggagc aacgaacggg 540

agccggagga gcgcgacact gcacgcagga gagcagacgg gaggggagac agcgcgggga 600

<210> 49
<211> 1098
<212> DNA
<213> Homo sapien

<400> 49
aacctcttca acaataaatt gctctttggg gacattttat gcacagaact gtgcaccctc 60

ctcagaacag caggtcttta atggcccatg tgatgagaag ggccccatca aggcagcagg 120

aatgggccac tctcccacac cccatgggcc aggccactgc cactcctgct gccctgcatc 180

```
cccaggttta tggctgcatg gtagaagtca cttctgtaag aaattcacct ttctaaaata    240
aagtatgctc tttttctga gacatctata gaataacttg tggcagagtg ttttaaaaac    300
tgatttggat tttttttatc ctttaaccgt gtgaaaggat ggaagggatt ttaggtggaa    360
gagaagttaa gaacagaaag atagagcagg tttttagagt gggagaatta atcccaaaga    420
aaaagagggc atggaaacaa atgtggatgc catgggctct gtgccagact tgccagtgct    480
gactggaaca ggccgggctc ctcactcagc ggctcctgcc tcagctgtgg ttcccgcagc    540
ctctgggtct cacggaaccc ttccttggga gttccatttc tcatggcttt gctcatcttc    600
cggcttcagg ctctgacttc atctcaggat gggatcggtg tgtgtctgtt ttcatagatc    660
cactacatca gaagtatctt tacatctctg tatctttaca tcccaaggtc aaggccctgg    720
caacctcaga ggttcccata gcttcagtct tccccaaacc atgccacttc ctcccatttc    780
tttgggtcag gaatctggct tttgttttcc atatttcttt ttcccaagac attgggaggc    840
atctggtgaa caacaccaat aaaacagttc tctccccacg gtcatccagg tcacttctct    900
aactcattcc tgcacacaca gcacacgtgg aatttgcctg tttagtctat gttcttgact    960
tgatcacaga cgcctgtaca ataaagcccc ttttcaacaa ggtgctgcag aatgataatg   1020
ctttccccaa aatctgaaac tgatttgtat cattgaagtt tttttctgta ttaaaaataa   1080
agcaaaatta aaaataaa                                                 1098
```

```
<210>  50
<211>  540
<212>  DNA
<213>  Homo sapien

<400>  50
ggtcgcggcc gaggtactcc cgcctcctgg agcggccgac cccacatgga ttctcaacag     60
gtggccggca catcttctga gcctcgctct ctcatctgaa agtggagtgt aagtccaaga    120
agattcattt agacaaagaa ggtggaaaaa aaggactttc tgggccagca agtcggatga    180
ccaccctcca aggggcagag gagggcccat tttgtgaaga agaaatcaac tacccggaaa    240
acgccacagg aggacatgtt tctgcagatg tagttgccct agaaacagaa gagtatgggg    300
gtgtgaatgt cttctctttt gggggcaaac actatgtcct tttcttttc tagatacagt    360
taattcctgg aaattttagc gagtttgttc ttgtggatat tttgaacaat aaagagtgaa    420
aatcaaaaaa aaaaaaaaaa aaaaaaaaaa accctgggcg gtacccatgg cgcaaagcct    480
ggtcccctgg ggggacactg ggttacccgg cccccaattc cccacaattg cggagcaacg    540
```

<210> 51
<211> 1028
<212> DNA
<213> Homo sapien

<400> 51
cggccgcggc atgaaaggcg gcgaggagag gcagcactgc tgctcttgac ttctgagcag 60

ggcttagaga gcctgccccg gcttaagccg agctgctggt gctgaccctg agcgccgagt 120

ccgcgagctc tgagtccgga gcctcccagc cgtggagccg tgggatgagg ggggcgttgg 180

gggacagggc aaagtcgatc ttggttgtac agccgcccga tcctagcgcg gagctgcgag 240

cctgaccggc cgcgtctggc atggtcagag aaagaatttt cttttcccaa ctccggcttt 300

tggttttgtg tgtccacctt gcgcaactcc ggagccagcc gaccccacat ggattctcaa 360

caggtggccg gcacatcttc tgagcctcgc tctctcatct gaaagtggag tgtaagtcca 420

agaagattca tttagacaaa gaaggtggaa aaaaaggact ttctgggcca gcaagtcgga 480

tgaccaccct ccaaggggca gaggagggcc cattttgtga agaagaaatc aactacccgg 540

aaaacgccac aggaggacat gtttctgcag atgtagttgc cctagaaaca gaagagtatg 600

ggggtgtgaa tgtcttctct tttgggggca aacactatgt ccttttcttt ttctagatac 660

agttaattcc tggaaatttt agcgagtttg ttcttgtgga tattttgaac aataaagagt 720

gaaaatcact ttggagtcac ttaatcttcg ttagaagggc agtttcttcc agggccattt 780

tctttcacca gatttgtttt tcctcgttcc caaatgaggt agttttaaaa atcaaagtcc 840

acttgctaac tcacctggga aagagactgc gacagaagga agagaagtaa atagacatca 900

ctctcaaact aaaagtgtaa ctttcattcc tggcagctga gattcagaac acaaagaaac 960

aaactcgttt acctttgagt atttccccccg tatgggtaat ttatctagag ctttcccaac 1020

aattaatc 1028

<210> 52
<211> 541
<212> DNA
<213> Homo sapien

<400> 52
acagattggt aaggtgacat tgtatcacaa agctagtctt tgagtccaaa gttttgtggt 60

tttatgttat gatatacttt tatcatggaa ttgtcttatt aaatgttttg ccagtggttc 120

ttaaagtgtg tttctgacac cagtagcatt gacttcactt agaaacctgt tagaaataca 180

aattatttgg ccccacccaa cacttgagtc acaaactttg cagatggggc tcaatctgtt 240

ttaacaagcg cttcatgtaa ttttgatgca ggcctaagtt tttgagccgc tgcagtatgc 300

atttctattt ttaagcaaag atcttggtct ttctttttgg acattgtaga aataacatga 360

acttgtcttt tgtttgtttt ggttttgttt tgttttaagc tcctgatctt tgttggttat 420

gttgcaaaag attgtatcag gagaagcctc agcatggaca ttggcatcct gacataaccc 480

ccattaattt agtattcttt ctgaaactca aatggattct caagtccaag agactatgga 540

a 541

<210> 53
<211> 261
<212> DNA
<213> Homo sapien

<400> 53
atgccatcag tggcacaggg ccctgtgccc tggcatctgg gttcacgctc tgctgttgct 60

gtcttcgaat tcctagtgat gtttgaacaa aggccctatg tttgcatttt gcactgggcc 120

ccacaaatca catggcccat cctgagaaga ggagtctcac acctccagtc tcctaaatca 180

cctctggaag tttttctcaa cgaaagaact gaagctttcc tcaaaagttc cgtaggggag 240

acagttcatc accataccca a 261

<210> 54
<211> 325
<212> DNA
<213> Homo sapien

<400> 54
gctctgtttt gtgttttgtt tggattgtgc tggttgtgtt ttgtgtttgt ggaaggtgtg 60

tgtgtgggtt tggcgagtac atgtcgcccg ggaccgctat ggctctgggt gcgcccacgc 120

tttttttttt tttttttttt tttttttttt ataatcaacc tataagggat ttatcaataa 180

ataaaccctt atttattata aggaattggc ttacacaata atggaggccg agaaggcccc 240

aagtctgctg tccgaaggtc tgagaaccag gagcactgat ggtgtcagtc ccagttcaag 300

ggcaggagaa gatgggtgtc ccagc 325

<210> 55
<211> 2461
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (356)..(393)
<223> a, c, g or t

<400> 55

gcctgaatag agctgtgcag cccaaggggt ggactgagcc agcagtggat atgcaccact 60
gagatctctt gctgtggaac gtaattgact ggggggggtcc ccgctactgc tctctgaatc 120
cattgataca gtcatgccaa ggctacattt cccatgggtt gtttccataa gaataacaat 180
aactgaatga agaaggtata ctaataatgc aggcctattc ctgtgaggta gggggctcct 240
ccaatgggcg actttggttt gagtgttctt catcagctga ccttaaactt tattggaatt 300
gtgctacagc ctaagctttc tgctactcaa cccgcctttc ttccctctct ccttcnnnnn 360
nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnaaaccta tggctctatc aatcaattaa 420
taaggattta taatcaacct ataaggattt atcaataaat aaacccttat ttattataag 480
gaattggctt acacaataat ggaggccgag aaggccccaa gtctgctgtc cgaaggtctg 540
agaaccagga gcactgatgg tgtcagtccc agttcaaggg caggagaaga tgggtgtccc 600
agcgccacag tcaggcagaa aattcaagct tcctccacct attttatttg ggtccttaga 660
agactggatc aagcccatcc acactgggga ctgcaatctg ctttattcag tccatcaatt 720
caaatgcgaa tttcctccag aaaaagcttc atgaacacaa ccagaaataa tgttcgatca 780
tatatggggc atcctgtggc ccagcgaagt tgacatgcat aaaattaacc atcacacctc 840
catggaggcc agaaatttta tgtattctga ggactctgtc tctctggctt cctctccatt 900
tcaaccctca ggcattcttt ttttccccaa ataaatgtct tgtaaatcta tttttttttt 960
acatgtgttt tttttggagt actggaacta ataaatttgg actttgttct gcaggtagtt 1020
tagttactta cagatcagct ttatactatt gaggcttgtt tttaagcttt ttgtgggggt 1080
gaggagcaag tgtagtgtag tttttagtca ggatagttta gccctacgta tgtaacacac 1140
aacctttctg ggatctttag cgaatgcctg gttttcagtc aggactctcc actgtggctg 1200
gatggaaatt agatatttcc caatctcacg ctaattttga gaattgttag gcttagtgtt 1260
ccgcaatgta cacaatgggg tctctacata gatttctgga ttttttgtc taccttcctc 1320
cttgttgata ctttgtcctg aaaattcaag ctatctcagt atcaactcca atcttcctct 1380
cctcaactca gcaagactct gatctcctgc gatgggatcc agagtgctca cctcagaaag 1440
gcagggcgat cggaggtctc actttgtttc cctgcctcag cgatcaaatt gtataaactt 1500
tttccctggc ttctggaaac actgatagcc aacctttcgg taaaatttct aacatatgta 1560
cttagagctt taatatgcta agtatttaaa tgctaagtat tttctttaaa attcatttca 1620
aaatattttt gcaattttgc cgtgatttat tttttggctc atgggtcatg aggtgtatgt 1680
tccttaaatt aatgcatttt ggaaattttt tgttaccttt atgtacttga tttctggttt 1740
aattcatact ctatatgatt tcaatctttt gaaatttctt gagacttgat ttgtgatcca 1800

gcctaacatg cgccccagaa cacatggtaa atgttttggg gtatacttta aaagaacatg 1860

tattctgatg tttttgaatg taatatctta tgtcttattc atatttatag taccaataca 1920

cagcacatag tagaaactta acatatattg agttaaataa ttcaaaggtt ttatccgatt 1980

ggtggcaatt caagaccaaa taagagagga tgatgatgac atcactattt ctgttaagac 2040

agggcctcat acaacataca ggaatgtcca gttgtcaagt catgcagttt ctcccctatt 2100

ctaaccaatg ttaatgccaa tactttgtga tgaaaattat cccagtattt ttcctcctat 2160

gtttccacca gttttccctt cattgattgt taagatttat atcactagag ctatttgaca 2220

gtaggaaaca attaccttag gaaaagttgg tgacattggt ctataaaggt cacggagaca 2280

taagaaatgg ttattttttc atttttcacc aaacaattca cgattgtttc taagattaca 2340

aaagattaga cgatagctaa tatttctatg caatggtcaa atttttcaag tagaatcatt 2400

tttaaatttt ccaagttcca atgtcacttt ctccttgaac acgactcaag gtcaaaactt 2460

a 2461

<210> 56
<211> 643
<212> DNA
<213> Homo sapien

<400> 56
ccgcccgggc aggtacacat gagtgcgtgt atgcccccag gctgggtcag ctcttctgtg 60

gattgcatgg cgtgtgatta aaagcccatg tgttcccaca catccacatc atgggaaggt 120

taatgtgtgc ctccttggaa ctgggtgttg gtgtccatgg aacttcctct ctgtatctca 180

ggtcagtagg cgcagaaacg cctcatgatg aagattcttg agccccattt ccaagacccc 240

tcacatccaa tcctgtcctg taacatccat caaggatttc cataggggtg actggtgccc 300

acccaagact gcaccagtgc ctgctcattg aggagagtaa ctgctggcca ggcagaaaga 360

atatgggctc tgcaatgaga cagacctgga ggggactctc ccgttgagca ctagcagctg 420

gaggagttgg gagttcatgg ctatcatggt tgtgttcaat cgattgtggg gatgacatgt 480

cattgtgtat ggaaggcggg gctcatggct gattggccaa taaaatggcg gctgccgttg 540

tcattgaaaa aacacaccac accacaacca aaaccgctgg ggcacacccg ggcacaaggc 600

cccccgggga aacgggttcc ccgcccaaat tctccaaatt aga 643

<210> 57
<211> 1611
<212> DNA
<213> Homo sapien

<400> 57
ctcctcccga ggaaccagtg gtgacagctg aggccatgtg agtaggatcc tgaatgaggc 60

tttatctctg gctgttcgtc ccatcgtcca ccgtggcacc agctccctca gccagccggg 120

atgggaccag cgactgagag agccagaggc agagaggtga gggtgaccat atcctggact 180

gtgagaggaa tgggactctg ggcctgtagc tgccaagcag gtggcaggtg ctccaggctg 240

tgatctgcac cctctgaccc ctgacattga cctcctaccc tgacccctgc ctgaccaagc 300

catgtctgaa caggaggctc aagccccagg gggccggggg ctgcccccgg acatgctggc 360

agagcaggtg gagctgtggt ggtcccagca gccgcggcgc tcggcgctct gcttcgtcgt 420

ggccgtgggc ctcgtggcag gctgtggcgc gggcggcgtg gcactgctgt caaccaccag 480

cagccgctca ggtgaatggc ggctagcaac gggcactgtg ctctgtttgc tggctctgct 540

ggttctggtg aaacagctga tgagctcggc tgtgcaggac atgaactgca tccgccaggc 600

ccaccatgtg gccctgctgc gcagtggtgg aggggccgac gccctcgtgg tgctgctcag 660

tggcctcgtg ctgctggtca ccggcctgac cctggccggg ctggccgccg cccctgcccc 720

tgctcggccg ctggccgcca tgctgtctgt gggcattgct ctggctgcct tgggctcgct 780

tttgctgctg ggcctgctgc tgtatcaagt gggtgtgagc ggacactgcc cctccatctg 840

tatggccact ccctccaccc acagtggcca tggcggccat ggcagcatct tcagcatctc 900

aggacagttg tctgctggcc ggcgtcacga gaccacatcc agcattgcca gcctcatctg 960

acggagccag agccgtcctt cttctcacag cggcctcagc gtccccagag ccgagccagg 1020

gtgtgagtgc atgtgaacgt tgagtacaca tgagtgcgtg tatgcccccca ggctgggtca 1080

gctcttctgt ggattgcatg gcgtgtgatt aaaagcccat gtgttcccac acatccacat 1140

catgggaagg ttaatgtgtg cctccttgga actgggtgtt ggtgtccatg gaacttcctc 1200

tctgtatctc aggtcagtag gcgcagaaac gcctcatgat gaagattctt gagccccatt 1260

tccaagaccc ctcacatcca atcctgtcct gtaacatcca tcaaggattt ccataggggt 1320

gactggtgcc cacccaagac tgcaccagtg cctgctcatt gaggagagta actgctggcc 1380

aggcagaaag aatatgggct ctgcaatgag acagacctgg aggggactct cccgttgagc 1440

actagcagct ggaggagttg ggagttcatg gctatcatgg ttgtgttaat cgattgtggg 1500

gatgaaatgt cattgtgtat ggaaggcggg gctcatggct gattggcaat aaaatggcgg 1560

ctgccgttgt cattgtctcc aaaaaaaaaa aaaaaaaaaa aaaccgcgga c 1611

<210> 58
<211> 617

<212> DNA
<213> Homo sapien

<400> 58
actgtgaagt cttcaggctc ttagaaggct ccagcctgag agagcccttt attattgcca 60
ttcctgtcct tcctcaaggc ctggtgacct gtgacctttc gctctgggca gggcccaggt 120
agatgggccg tcatccgggc ctgtaagccg tactatgatt tctgcattga tttacatatt 180
ttttactgtg atcttggttc caaacacaga atcgtcaccc cattctccct tgaatgtgcc 240
ggatccttgt aaattctcat tcacctactt gttcttaggt gtgtatgtgt gtgcgaaact 300
ctatgttcaa gaaagaaatc atacaaagag taacgaacca tggttctgtt ggccattgga 360
cgaaacttgg tttttggact ttcttaccta acattaattt tgctcttgcc tcggtttaca 420
cacacacaca cactacaaca aacacaacac aaacaacgtt ctgggccaac accacgcggc 480
gccagcgccg gctccctggg ttgaaacttg gatctcttcc cgcgccacaa ttctcccaac 540
aactataatg agcacaagga ccacaaccat acacaagaac aacacaaacc agcgacacaa 600
cagagacaac acacaac 617

<210> 59
<211> 913
<212> DNA
<213> Homo sapien

<400> 59
caaaaccaca cccatgcaca cacataccct cagcccccac acacaccccg ttgaacccgt 60
gagtctatca gggcatccta aaactccgtg agttgacatt tcagtaattt caggggaagg 120
tgttttccag ggatggggtc tcccaggttc agatagtgcc tttggctgca aatgctcctt 180
tagctaaact tttcctcagg aagaattcat tattctagac attatgtgat atatctgtta 240
ggaataaaag gtgcttaacc ttcctccctg ggatgtggga gaaggtgctg gaggttgtac 300
tgtgaagtct tcaggctctt agaaggctcc agcctgagag agccctttat tattgacatt 360
cctgtccttc ctcaaggcct ggtgacctgt gacctttcgc tctgggcagg gcccaggtag 420
atgggccgtc atccgggcct gtaagccgta cttgatttct gcattgattt acatattttt 480
tactgtgatc ttggttccaa acacagaatc gtcaccccat tctcccttga atgtgccgga 540
tccttgtaaa ttctcattta cctacttgtt cttagtgtgt atgtgtgtgc gaaactctat 600
gttcaagaaa gaaatcatac aaagagtaac gaaccatggt tctgttggcc attggacgaa 660
acttggtttt tggactttct tacctaacat taattttgct cttgcctcgg tttacacaca 720
cacacacact acaacaaaca caacacaaac aacgttctgg gccaacacca cgcggcgcca 780

gcgccggctc cctgggttga aacttggatc tcttcccgcg ccacaattct cccaacaact 840

ataatgagca caaggaccac aaccatacac aagaacaaca caaaccagcg acacaacaga 900

gacaacacac aac 913

<210> 60
<211> 554
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (304)..(430)
<223> a, c, g or t

<400> 60
tggaaaataa agtttaaaac cagattgccc agagcaagac tctaatgttc ccaacggtga 60

tgacatctag ggcagaatgc tgccattttg aggggcaggg ggtcagctga tttctcatca 120

agataataat gtatggtttt tacactaagc aactgataaa tggacaattt atcactggac 180

aatctccctc tgcttcttta atggggccag ctttgcagcc ctgcagcctg ggtagtcgca 240

cacatttcca tgcatccaag gcccccgtgc ttgggagaat gatctgctag tgccatttta 300

aatnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 360

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 420

nnnnnnnnnn tcactgtgtc cggcataaag tagaacattc ttacaagaaa taaatatttc 480

gtagtcatgg agaagaacgc gaaaaaaaaa aaaacaaaaa aaaggctggg ggtaaccagg 540

gcccaagcgg ttcc 554

<210> 61
<211> 1401
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (803)..(929)
<223> a, c, g or t

<400> 61
aattattttg ggtctctgtt caaatgagtt tggagaatgc ttgacttgtt ggtctgtgtg 60

aatgtgtata tatatatata cctgaataca ggaacatcgg agacctattc actcccacac 120

actctgctat agtttgcgtg cttttgtgga cacccctcat gaacaggctg gcgctctagg 180

acgctctgtg ttcactgatg atgaagaaac ctagaactcc aagcctgttt gtaaacacac 240

taaacacagt ggcctagata gaaactgtat cgtagtttaa aatctgcctc gcgggatgtt 300

actaaactcg ctaatagttt aaaggttact tacaatagag caagttggac aattttgtgg 360

tgttggggaa atgttagggc aaggcctaga ggttcatttt gaatcttggt ttgtgacttt 420

agggtagtta gaaactttct acttaatgta cctttaaaat agtccatttt ctatgttttg 480

tataatctga aactgtacat ggaaaataaa gtttaaaacc agattgccca gagcaagact 540

ctaatgttcc caacggtgat gacatctagg gcagaatgct gccattttga ggggcagggg 600

gtcagctgat ttctcatcaa gataataatg tatggttttt acactaagca actgataaat 660

ggacaattta tcactggaca atctccctct gcttctttaa tggggccagc tttgcagccc 720

tgcagcctgg gtagtcgcac acatttccat gcatccaagg cccccatgct tgggagaatg 780

atctgctagt gccattttaa atnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 840

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn nnnnnnnnnn 900

nnnnnnnnnn nnnnnnnnnn nnnnnnnnnt cactgtgtcc ggcataaagt agaacattct 960

tacaagaaat aaatatttcg tagtcatgga gaagaacgct cctaaaatga tgaacgcacg 1020

acggaaaaga gtagggaaca ttttgcttga tgagaaaatc cgccagcaag gatgttgggc 1080

tctaagcaga actgaagctc tggaattaag aacacagcca aggaagagct ctggactctg 1140

agtttaaaga agctgactga cttgtaaggc aattccaggt aagattggtg aatcaagtta 1200

agaatcaaaa gcaactgaga tcaacgtgga ggcctggaag gtaagggcca tattttacct 1260

agatactagc ttagagactt gctacattgg cactgtattt taagtatgtt atttagtagt 1320

attgtgaaat caactggttt caacattgaa aaggataaaa atagcttatg aaaacaaaac 1380

ggtttttttt tttttttaaa a 1401

<210> 62
<211> 568
<212> DNA
<213> Homo sapien

<400> 62
agatgctgcc gagcggcgca gtgtgatgga tagtccaaaa aaaaaaagta ttaaaatgtg 60

attgatgtaa tttaccatgt ttactttatg catgcatttt attggggagg ggacgtgtca 120

gaataataca cccaaatcta gtggtctaat ttcatagtgc taatctggtt tatattggca 180

ttaaacgata ctgcgaagga gctagatcat tttacaagag ttgtaggttt gtcttatgtt 240

ggaaaagcag tcctctatta atatcatgtg tgaagagtat ctgttcacaa gatttatgag 300

attatgacgt gtttcagaga atgtctacta gtatatcttt acagtatttg cctgttgaac 360

tccctgcaca aactggaatt actttccaga agacttaggg aatgcaaata tgttactcat 420

aagatgcatt ggagtatggt aaataaaaca aaccattttg gattggttta aattggctcg 480

ttacagttct cttgtgggga gggactttgt cagtcatttt ggcatcttaa gctagactaa 540

actttttgtt gttgttttcc taaaacca 568

<210> 63
<211> 791
<212> DNA
<213> Homo sapien

<400> 63
tggtctatgg taattttta tagcagtccc agccaagaca gtgcgctcat ttactacata 60

ccatttatat tattatatag gctcctttca gaaacccatg ttcaaataag agataagata 120

ctgaaacaca taacaccttc actagttttt agtatacaaa tattgagaaa tagttgttat 180

taactatctc atccaagaaa tgcagattca tgttgtttct aaatttttta tatatattga 240

ccaaaatgaa gaaacttaac accatcctag attttagctg cccaaagaat gaaaagaatg 300

aaaaaaaaat cttgtgaaaa cccacaagtg atatggatct aatttatggt taaatagata 360

tagataacaa acagaatacg cctgtttaaa actgttaaaa tgacattggt tctaattata 420

cttttattta aattgaaaga caaggcattt atatggtatc tctaaccatc acaactttgg 480

tgtgacaaaa agaaattatc accaaaatac acctccttaa gtaagtgtct gatttcacac 540

ttccagaaaa agtgctcttt ctggtcaagg ccagcaagaa ttgagaaaga ttaagaaagt 600

gcttcaaaga tgtttattac aaagttgtca taaaaactgt gaagtagatg tagacatcaa 660

gcataccaaa taaagtaaaa actgtcctcc ggcaaaacaa caacccaaaa aaaaaagcgg 720

ggggggaccc ggggccaaaa cgggtcccgg ggggaatggt tccgccaatc accccaacaa 780

aaaaaaaagg a 791

<210> 64
<211> 1523
<212> DNA
<213> Homo sapien

<400> 64
gggagatgct gccacctagg ttacttgtag gaccctatac ggcaacctcc tttgccagga 60

actatttata aacatcctgc aggaaaatgc agtgaagtag aagagacagg gatatcccag 120

aaggttatgc aaaacatcaa gagaagatga gaggagtcta tatgtcagaa tacacatttc 180

ccaccttgcc caacagtaga aaaacataag aagagaaaaa cattaaaaaa tgacaaggaa 240

gttaatggaa gtcagcaatg tgatggtgtt tggaggtgga gccttcagaa ggtaattaat 300

gcccttgtaa gaagaggcca gagagcttgc gcaccttctt cctgccatgt gaggagccaa 360

gaagccggct gtctgcaacc tgcaagagga ccctcactag aagctagcca tactggcatc 420

ctcatcttgg ctttccaact tccagaactg tgagaagtat atgttgtggt ttagtcaatg 480

gtctatggta atttttttat agcagtccca gccaagacag tgcctcattt actacatacc 540

atttatatta ttatataggc tcctttcaga aacccatgtt caaataagag ataagatact 600

gaaacacata acaccttcac tagtttttag tatacaaata ttgagaaata gtttgttatt 660

aactatctca tccaagaaat gcagattcat gttgtttcta attttttata tataattgac 720

aaaatgaaga aacttaacac catcctagat tttagctgcc caaagaatga aaagaatgaa 780

aaaaaaatct ttgaaaaccc acaagtgata tggatctaat ttatggttaa atagatatag 840

ataacaaaca gaatacgcct gtttaaaact gttaaaatga cattggttct aattatactt 900

ttatttaaat tgaaagacaa ggcatttata tggtatctct aaccatcaca acttttgtgt 960

gacaaaaaga aattatcacc aaaatacacc tccttaagta agtgtctgat ttcacacttc 1020

cagaaaaagt gctctttctg gtcaagccag caagaattga gaaagattaa gaaagtgctt 1080

caaagatgtt tattaaaaag ttgtcataaa aatgtgaagt agatgtagca tcaagcatac 1140

caaataaagt aaaactgtca tcaagaagat tcaacagcta tgaaaagagt tcttcaaaat 1200

atgatatgtt tttctagatg ataataaaat ttatcaattc caaatgtcca cattagtctt 1260

tcataaagac accaatgagt cacaggaaaa aaattaaaaa taaaaaaacc ctatctcagg 1320

gaatcatgct aacaacctga tgtgttttct tccacatatt tatgtctgct tataagtatt 1380

tacaaacata tattcgcata tatgcatttt gaatttttttc tgttgctgca cttaaatttt 1440

tttcataata aaacaagact cctgcaattt gctttttttag gtagactatg tatccctgac 1500

aaccatccag gtcagcttga tga 1523

<210> 65
<211> 377
<212> DNA
<213> Homo sapien

<400> 65
ggtcgcggcc gaggtacaaa agtgcaaaca aggttagtga ttaacaactt accatcaata 60

taccacttca acatacttta cattcagcca aatactgaag gtttcaccgt ggaaaaacac 120

ttttatcact tttaaagtaa cttgactatg ttcaccctga gtgctcttgc ctcagtatgg 180

caactgatta tgagttcagg ttaagagcaa caccagggaa tacagaaacc cacgttaagt 240

tggccattct gacatgaatc tatacttgaa aatgaaaaca atcccaaaga aaacctgtat 300

gtcaaaaaca gaactgttcc tgcctttcac cccaaaatat ttaaaactaa atctaagcca 360

cttttaaaat gcatgct 377

<210> 66
<211> 1703
<212> DNA
<213> Homo sapien

<400> 66
ccaggctgga gtgcagtggt gtgatctcca ctcactgcaa cctccacctc ccagcttcaa 60

gtgattctcc tgcctcaacc ttccaagtag cttggattac aggcgtgcgc caccacagct 120

ggctaatatt tgtattgtta gtagagacag ggtttcacca gtgttgtcca ggcttgtcga 180

acttctgacc tcacgtgatc cacctgcctc agcctcccaa agtgctagat tataggcgtg 240

aaccactgcg cccggccagc atgcatttta aaagtggctt agatttagtt ttaaatattt 300

tggggtgaaa ggcaggaaca gttctgtttt tgaaatacag gttttctttg ggattgtttt 360

cattttcaag tatagattca tgtcagaatg gccaacttaa cgtgggtttc tgtattccct 420

ggtgttgctc ttaacctgaa ctcataatca gttgccatac tgaggcaaga gcactcaggg 480

tgaacatagt caagttactt taaaagtgat aaaagtgttt ttccatggtg aaaccttcag 540

tatttggctg aatgtaaagt atgttgaagt ggtatattga tggtaagttg ttaatcacta 600

accttgtttg cacttttgta caccactgct tgcactagga tcttggtgtg aattttcaat 660

tgttttacag tgtatacaga ttattaagga taatttatat aaagatgttt ctgtttaact 720

ttgtgtgttt tacaacaaag agctataata gatggttaaa cgtttttgaa ttgtgtttat 780

atgttagttt gattagtatt ttatttttcc cttcctaaca ctcaaattca tggcaggtga 840

aaagataata gaacataatc aaactaacat ataaacacaa ttcaaaaacg tttaaccatc 900

tattatagct ctttgttgta aaacacacag agttaaacag aaacatcttt atataaatta 960

tccttaataa tctagtatac actgtaaaac aattgaaaat tcacaccaag atcctagtgc 1020

aagcagtggt gtacaaaagt gcaaacaagg ttagtgatta acaacttacc atcaatatac 1080

cacttcaaca tactttacat tcagccaaat actgaaggtt tcaccatgga aaaacacttt 1140

tatcactttt aaagtaactt gactatgttc accctgagtg ctcttgcctc agtatggcaa 1200

ctgattatga gttcaggtta agagcaacac cagggaatac agaaacccac gttaagttgg 1260

ccattctgac atgaatctat acttgaaaat gaaaacaatc ccaaagaaaa cctgtatgtc 1320

aaaaacagaa ctgttcctgc ctttcacccc aaaatattta aaactaaatc taagccactt 1380

ttaaaatgca tgctggccgg gcgcagtggt tcacgcctat aatctagcac tttgggaggc 1440

tgaggcaggt ggatcacgtg aggtcagaag ttcgacaagc ctggacaaca tggtgaaacc 1500

ctgtctctac taacaataca aatattagcc agctgtggtg cgcacgcctg taatccaagc 1560

tacttggaag gttggtgagg cacgagaatc gcttgaacct gggaagcaga ggttgcagtg 1620

agtggagatc acaccactgc actccagcct gggtgacaaa gcaagactcc atctcaaaaa 1680

aaaaaaaaaa aaatgagcgg tcg 1703

<210> 67
<211> 456
<212> DNA
<213> Homo sapien

<400> 67
atctctttaa ataattagca agaagggaga caagatgcag gagttcactt ggctctttga 60

aaaggaaaac tttaaagtca gtggttggac tgagtcccat gaagccagat cacttctgac 120

tgcaaggagc ttggaaaagc aagtatctgg atctttttacc agctaaattg ggaggaacta 180

taaaatgaga aaagattgat gaatattaag tagaagagtg agatggtcat ctttgcattt 240

aaaaaagatc atttgctgta gttgtatgga aaatgaattg gagcaggcga tgaggcttcc 300

tctttgaaga tcacaggtga gaagattagg tgctttctca gaagcccagc aacctgatgg 360

gagtgtggag tgagcaagac ccaaatcgga gcttcatccc tgcatggttc attttgctta 420

tttggcaaac ttgccctgca gaatctactc aagctt 456

<210> 68
<211> 380
<212> DNA
<213> Homo sapien

<400> 68
ccgcccgggc aggtagaggt tgtagtgagc cgggatcacg ccactgcact ctaggcctgg 60

gcaacagaga gagagactgt ctaaaaaagg aaaagaaaaa aatttatacg ccaaaaaaga 120

tattctgaga taacctgtag ttaccactaa ctttgtgaca aaattataaa aatccacagc 180

catctatgaa tctgtaggca gacctgaagt ttgaacgact ggtgaagaca tctgcatttt 240

ctttatagcc aagttaggat aacaaaaatg caaacaagtc attaatattt actatatgca 300

agatacagaa acgatgaacg gaaggagtaa gaagttatcc ttcgtggaac tatttaaagc 360

aaaaatgcaa aataccaggg 380

<210> 69
<211> 2177

<212> DNA
<213> Homo sapien

<400> 69
ttccaacatc tcatttctcc catgaactat ttggaaaaag ctgcaggcgt aatattggat 60

ccctaaatac tttattctcc ttataccatt atcagaccca agtatcatct aatagtccat 120

aatcaaactg cctaaacagt ttctacactg tctttttaac tatttcaaac tatcaaggtc 180

cgcattttct tccttagaac ttttagtctt tttcttcccc aaaatatttg agtccatgcc 240

agttgccttt agttgtaccc aaataatggt ttgtctattt cctaaaagta gtactcttaa 300

atttaaattt agtgttattt ttgttgtcat cgttccttct tcctcatgtg gttgtgcagg 360

cagagcttga gcatccagat ttcaaaatta aaaaataaaa gataatctag tttaatatat 420

agtagttgaa tcaccttaag tctagactgc tgtatgagca cccattatct ttcactatat 480

tccatcatcc ccctccccca tgaactattt ggaaaaagct gcaggcgtaa tattggatcc 540

ctaaatactt tattctcctt ataccattat cagacccaag tatcatctaa tagtccataa 600

tcaaactgcc taagcagttt ctacactgtc tttttaacta tttcaaacta tcaaggttcg 660

cattttcttc cttagaactt ttagtctttt tcttccccaa aatatttgag tccatgccag 720

ttgcctttag ttgtacccaa ataatggttt gtctatttcc taaaagtagt actcttaaat 780

ttaaatttag tgttattttt gttgtcattg ttccttcttc ctcatgtggt tgtgcaggca 840

gagcttgagc atccagattt caaaattaaa aaataaaaga taatctagtt taatatatag 900

tagttgaatc accttaagtc tagactgctg tatgagcacc cattatcttt cactatattc 960

catcatcccc caacatatcc acagtagatg aagggcagtt tgctcaaaca ttgttttgat 1020

cctgtcatgt ctgttcagaa atgcctgtct attcagaaac ccacgtctaa taacaaaatc 1080

ttggactggt tactatcaaa acccaacaac atacagactc ctcagctagg ccctagggat 1140

atttttctac cttgatttcc aaatgttcat tgaaagaatg cttaattcta atttggaaaa 1200

aagtttttgg cttcccactt ctgctttaca cgttcatctt tcttgaaatc aaatccaatc 1260

caatctatat tctaagaacc tgctcaaatc ttggttcttc aaagctttcc ctggtatttt 1320

gcatttttgc tttgaatagt tccacgaagg ataacttctt actccttcct tcatctttct 1380

gtatcttgca tatagtaaat attaatgact tgtttgcatt ttgttatcct aacttggcta 1440

taaagaaaat cagatgtctt caccagtcgt tcaaacttca ggtctgccta cagattcata 1500

gatggctgtg gatttttata attttgtcac aaagttagtg gtaactacag gttatctcag 1560

aatatctttt ttggcgtata aattttttc ttttcctttt ttagacagtc tctctctctg 1620

tcgcccaggc tagagtgcag tggcgtgatc ccggctcact acaacctctg cctcctgggt 1680

tcaagagatt cttaggcctc agcctcccga gtagctaggg ttacaggcgc gcaccacctc 1740

catgcccagc tcttttgtat ttttaagtag agacagggtt tcaccatgtt ggtcaggctg 1800

gtctcgaact tctgacttca ggcaatccgg ccgcctcggc ctcccaaagt gctgggatta 1860

caggcacaag ccactgcacc cagccttatt accataaatc atcttgatgc tggtacctga 1920

taagattcta tttgcttttc tttattcata gagaccacaa acagatcgca gatccaggtt 1980

tctcaaactg gagcatctgc ttaattttcc cataaaatca gtcttattct ttctgacagc 2040

tctgagactc ctccggccac gactaggtgc tgtcctggag gaaacggtgg aggacggccg 2100

cacaaaaacc aatctacctg atgaaaactc cgttcccttc tcgccagaaa cataaaatgc 2160

gatggagacg ctcgtgc 2177

<210> 70
<211> 226
<212> DNA
<213> Homo sapien

<400> 70
tctcatgccc attcaatatg gaatgttctt cgcttgctga atttaagcct gtattttaag 60

gttttgtggt tcctcggcca caatgggtga tgtcactgat agaacgaagc tgagtttcta 120

agggtttggg gctgtgcaag agtaaacact agagcttgag ttgttatcca gctggcaagc 180

acggaagtct ttgaagaatg taatgtaaaa agggaacaag aatgta 226

<210> 71
<211> 2554
<212> DNA
<213> Homo sapien

<400> 71
gcgggagagc cctgtcctta aacacattag gacaagtagt taaaacaggg ccaagaagta 60

tggctgtgta gtgatcactg tacaagcaca cctggctgaa taaaccagtg ggggataaaa 120

tccagctcac ctgccgctgg ctatgctttg tgcctcagga caagggtgtg cttccttgct 180

aattgacagg aaccatcttc ctgcccaact gcattcccac tgcgtaggca ccttatctgc 240

ccaatggggc tgtgaaccct aattggaagc tttgcaattc ttaacactat atcttcttga 300

gctgggtttg agtccctatc caatcaagat gaaggcctga gaggactact caagttctaa 360

catgatgtgg gggcaaggca tagtagtcca gatccgggac atgaggcagc ttttggctta 420

gtatgacaat ctaatagttc ctaaaataga attatcccag gatggagctc cgtatgacag 480

aagggctctt cataggtagt tggtaggggg aattgtgtat catgtaagaa gtaggaccag 540

```
atgtctttaa aaagaccttc caactctaat gctacatgag tctgtctagt tgttatgttc    600
caacagggac agctcttaaa atagtgtggc aaagcaagag atgagatttc cagtgctgac    660
tcggtggtgg aatgacttta gggcaggtat ttaacctcca cttccccaag tacacaagtt    720
atttcacaac tcttggcaaa aacagtgctg taaaaatcgt aagtttattt gttaaaaaaa    780
atactgtatt tgaaaagtac cttccttctg ggattttcaa ataatttgta cactacattt    840
tattcatcta cacattggaa atgagtaaac tggtgaacat atagcttttt atacatttaa    900
cacaaccagt gcaaattctc ctgcctctga gaaggcagag aagcccttta ctcagaaggt    960
cttcaattct agcattactc caactcctag ggaaatttcg ggtgggtgcc tatggctgta   1020
tgaccatctg attcctcagg gacaggacag gaattcagca agggagctta aaatatttta   1080
agtaattgtc aacattccat ggtgactctc cccaaaaatc tagtggtagg aaaataatct   1140
gtacttattc ctctttctgc acacaaagcc ctcatttaaa tttgtgagcc tgcttgggat   1200
ccattaccta gccattcaga gatcctgtca aatgcacagc agattggata ctcaccatcc   1260
caaaggggtt cctcccacct ggatggggcc aatctctagt tgacagtgcc cctcagagtg   1320
caccatggag atggaatgtc ccttccagag agacttttac acagggaaaa gcatttgttg   1380
gctgggctcc aactctcatt tggtacaaaa agctttacat tcttttccct ttttacatta   1440
cattcttcaa agacttccgt gcttgccagc tggataacaa ctcaagctct agtgtttact   1500
cttgcacagc cccaaaccct tggaaactca gcttcgttct atcagtgaca tcacccattg   1560
tggccgagga accacaaaac cttaaaatac aggcttaaat tcagcaagcg aagaacattc   1620
catattgaat gggcatgaga tatgcctatc agattgtgtg tgtgtgcgcg ttttttaaag   1680
acagccaatt acatcgtatc tagtcaaatg agcggattct aaagcagcct gctgggatgt   1740
tccacttagt ctaatgctgt tgccactgta cgccacagca ccggacagtg ttctttggga   1800
catctctggg aaatgctctg gaacatgctc cttgatggaa aacactaatt tttgaaagaa   1860
gtagatgtct ggaggcaggt ctggtgaata aactgaatag tactgccttg gaccccagct   1920
gaggggtggc agtaagcaat gaggatgggc tataaagctg ttaactggct aagggccatc   1980
cttgggcagg catttcagac acatctgtag agagggcagt agcatctccg ataggccagc   2040
tctgaaggaa gcttaatgct taatacagtc acactgcata aattagctta gaatgctctc   2100
ttgggtaaaa aatattaata gtgtatatgc acttgaaaag caaaattcct caagaaaaaa   2160
agtttaatag caaggagttt ccatcagtcc cggtctttgt gaggattacc acaacaaaca   2220
cttaaaagga tacaacaggt acttattaaa tgctgccttg ccttttacct cttccttttt   2280
tttttttttt tgagatggag tctcgctctg ctgcccagcc tgaagtgcag tggtgtgatc   2340
```

tcggctcact gcaacctccg ccttccaggt ttaggtgatt ctcttgcctc ggcctcccga 2400

gtagctggga tggactacag gcacatgtca ccatgcccag ctaatttttt gtatttttag 2460

tagagacggg gtttctgtgt tagccaggac ggtctggatc tcctgatttc atgatccgcc 2520

cgcctcggcc tccctaccct cgtgccgaat tctt 2554

<210> 72
<211> 583
<212> DNA
<213> Homo sapien

<400> 72
cagatcatga agcaattatc ttcctggaag ggtttttagc tatgctctcc agttgcctca 60

gcagctttgg ctctgatgcc acagtgagcc caaggtggaa ggtgatggaa cagcatcaca 120

tctgcaggct cagtgtgtcg taaggtgagg gtaaggggag ggcaagtgta gacggatgaa 180

gaagatttct ccctattgct tccattttga tatttcttta acttcacatt tcatccatca 240

ttcctagaca gttgcctagt tatagaggat ttcttttatc ttttttatca gaggcatgcc 300

aggtggaagt gaggctgctg ctggcctaca actccagtgc tcgcattcca aaatgcccct 360

ggatggaggg tggtgagatg tcaacacagg tggaaaacag atccgagggc accataccca 420

tacagacaac ctgtaaaagt cataataaag ccccacactg cacggagcta aggcacaaac 480

aacgcttccc aaccgatggc taagggccaa ctaggcggca gatgagcaag ccgaagcatc 540

accgaaatga agcagctcag aagaggacct aagccccggg aca 583

<210> 73
<211> 981
<212> DNA
<213> Homo sapien

<400> 73
gaaagaatga gatgttttca gacattttag gtccctgaga catgttcctg ttcattggcc 60

agaaactttt tggcgaacca cttcctattc aaaggcttcc tctccactaa taaagtagtc 120

tgtggtacat gcagccctga ggcttgagat ggaactgcgc aggaagagcc caactgggta 180

tcagaataag ccacatgcac cttctgaaac tgcccaaatc cacacctgca taagaatttg 240

agcccagttc ataaagcaga tcatgaagca attatcttcc tggaagggtt tttagcttgc 300

tctccagttg cctcagcagc tttggctctg tgccacagtg agcccaaggg gaaggtgatg 360

gaacagcatc acatctgcag gctcagtgtt ttgtttggtg agggtaaggg gagggaatgt 420

agacggatga agaaatttct ccctactgct tccattttga tatttcttta acttcacatt 480

tcatcctcat tcctagcagt tgcctagtta tagaggattt cttttatctt tttttcagag 540

gcatgccagg tggaagtgag gctgctgctg gcctacaact ccagtgctcg cattccaaaa 600

tgcccctgga tggagggtgg tgagatgtca ccacaggtgg aaaccagcat cgagggcacc 660

attcccttca gcaagcctgt aaaagtttat ataatgccca aacctgcacg gcgctaaggc 720

aaaaacagtc ttcccaaccg tggcctagag ggcccttctt aggtgtcaga atgagccaag 780

cctgaagcac ttcacctgga attgatgtgt aggcttaagg agtatgtgac ccttacagtc 840

tcatctggta tcaaacacag gataaattgt ttcttcatta aaaaataaaa aaccttcaag 900

tctacttacc cttctcctgt ccacaataaa gttgagaaaa caaaaaaaaa aaaaaaaaaa 960

aagatcttta attaagcggc c 981

<210> 74
<211> 401
<212> DNA
<213> Homo sapien

<400> 74
gccgcccggg caggtaccag gcagagggag gagcaccaag gtgggggata tttaggggac 60

ctctttcctt caggaccaca cccttctagg tgaaagcacg aacacttgat tactttgcat 120

tccatctgca aaaacaaatt taggttttga atatggtgaa aaacgaagaa aggaaaatat 180

aaaactctgt attttatata cagtaaggaa taatggaggc tgataatgat cttgtgatca 240

gctaagacaa tgtcagtaag caggtgaggt agggtgcttt ctatgggcaa aagggtgaat 300

atcttgaatg accagaaatg actcgaagag ctgcattact atcatggtag catgcatgaa 360

gtgatacatc taaacctttg ctaacctaac attattactc t 401

<210> 75
<211> 1847
<212> DNA
<213> Homo sapien

<400> 75
gccgatcttt tttttttttt ttttttattt ataaatttat tgcctgtttt attataacaa 60

cattatactg tttatggttt aatacatatg gttcaaaatg tataatacat caagtagtac 120

agttttaaaa ttttatgctt aaaacaagtt ttgtgtaaaa aatgcagata cattttacat 180

ggcaaatcaa tttttaagtc atcctaaaga ttgatttttt tttgaaattt aaaaacacat 240

ttaatttcaa tttctctctt atataacctt tattactata gcatggtttc cactacagtt 300

taacaatgca gcaaaattcc catttcacgg taaattgggt tttaagcggc aaggttaaaa 360

tgctttgagg atcctgaata cacctttgaa cttcaaatga aggttatggt tgttaattta 420

acccctcatgc ataagcagag gcacaagtta gctgcatgtg ctctagactg tagagcgagc 480

caccgttgag aagcaaagga cagcagcagg aagagcaatg gaacctcctc aggacttacc 540

aggctgctgc acaggatcta gcttctccca cctaagatgg gcacattgaa agccttgttg 600

cagcagcacc cccatctgtg gaagcacagg ctgcctgcac ttctccagct gctctagcac 660

ctgacttcct ggtagtcagg gtaccaggga gagggaggag caccagggtg ggggatattt 720

aggggacctc tttccttcag gaccacaccc ttctaggtga aagcacaaac acttgattac 780

tttgcattcc atctgcaaaa acaaatttag gttttgaata tggtgaaaaa cgaagaaagg 840

aaaatataaa actctgtatt ttatatacag taaggaataa tggaggctga taatgatctt 900

gtgatcagct aagacaatgt cagtaagcag gtgaggtagg gtgctttcta tgggcaaaag 960

ggcgaatatc ttgaatgacc agaaatgact cgaagagctg cattactatc atggtagcat 1020

gcatgaagtg atacatctaa acctttgcta acctaacatt attactctca agctttatta 1080

tcctcaaggc ttaaatggct gtagctgttt aatttaaaag caaggcttaa aaaatagagg 1140

ttactcataa ttccctttcc atatcccttt ttgacttgaa aattatttca ccaactactt 1200

ttctggaatg ctgcttataa tacatattca cagattgccc tatgtgttat tctagtcatt 1260

ggcccgtttt gcttataaaa aaggccatgt tttgtattcc tacaaaatct gcagacattg 1320

ttaacataat acacgtcatt atacatcata tgtatgctac atctactcac tgacatttaa 1380

aaaatgagct attttcaaag actaacacag gatctgttac tgagacgtgt aggaaggagc 1440

tcagtgtaaa atattttctt tggatagatc ccttcaaagg gattaaaaca cacaaaatat 1500

tatttatact aaactttctt aaatgttcta tgatatttct atttcaaaat tctcttattg 1560

tgagaatatg tgaaatatag atgtagcaaa ttcaacacat aagcttatac cccttagctt 1620

gagtaaaaga cacatatatg gcttcccagc accaagaaga tggaagaaac tctactgcaa 1680

ctacttccct ttttccaagc agctcaaaat gctttagcaa ataccttgtg attctttttt 1740

ttttttttt ttttgagacg gagtctcgct ctgtcgccca ggccggactg cggactgcag 1800

tggcgcaatc tcggctcact gcaagccgcc ctcgtgccga attctat 1847

<210> 76
<211> 522
<212> DNA
<213> Homo sapien

<400> 76
attttactct agtattaatg tggttttata aatgattata tgccttatat tctgggggga 60

aagaaatgtg aaaatgtgct aacgtagaca gaaacagaat atataagtcg ttttgaatgt 120

tatttctttt ttaaaaaatt tgcttggtgt catatagcca aaactattca tggtgacagt 180

ttcattgcta tactttttat atgatttcag cgaattgaaa acatgtatat aatagcaaaa 240

aactggactt catgctgagt atagatgata catataaaag aagtcaaaat ttggagaaaa 300

aatttaaaaa gataagtaga aaaatgaagt aactgtagaa accatactta ctctttgatc 360

tcaaatgctc aaaaactgaa tgaaaatgtg aatttaggcc gaccaggtag tcttgtcaat 420

aaactaaaag caaaaacagg aaaattgaga aatatgttac aactataaca acacaaaaca 480

gcatagtttt gaaacacttg cagttcttaa atataaaagc tt 522

<210> 77
<211> 1643
<212> DNA
<213> Homo sapien

<400> 77
actgtcaatc atcaattgac attaacatgg tcaattaagt aatgtttctc acccaacttt 60

aaatttccat agtcataacc atggaaacat acaaaaaaca aacatgcaaa taaaatgtca 120

aaataattga gctgagtact ttgcatgctt taggaaataa gatgtagggt ggttctttgt 180

gccaatatat tcaagtaatt ggtttatctt cccatgtttt gctgctctaa acatgatcta 240

atataactct cattcatgtt gacatagcag agagctgcta ggagtaaacc tgttttctac 300

acattaatca agctgttctt tcaaagtatt gtttgacaca ttgaatgttt tttattctgg 360

aatattatca cagcaaaacc tcattaattg gatgctatca aaattatgaa aggaaatctg 420

agtgagcaca cttgttttga aaagaaattg gtaaatactt ctatgatgca gttttaagtt 480

atacaattaa ctgctatttg gaatttaata agtccactat aagcaatgtg cctgcacacc 540

aattaaaggt tggatctgtc tcttcttgac aattttttag aagccattat ttcgttacca 600

aataaacctg aagttaagaa atatttatat ttacatctat ttatatctgt tggagaatat 660

ttcataactc agacttggtt gttttacaca gacttctccc cattatccaa catagtgaga 720

tttttctata gttctatatt ttactctagt attaatgtgg ttttataaat gattatatgc 780

cttatattct ggggggaaag aaatgtgaaa atgtgctaag tagacagaaa cagaatatat 840

aagttgtttt gaatgttatt tcttttttaa aaaatttgct tggtgtcata tagccaaaac 900

tattcatggt gacagtttca ttgcttactt tttatatgat ttcagcgaat tgaaaacatg 960

tatataatag aaaaaactgg acttcatgct gagtatagat gatacatata aaagaagtca 1020

aaatttggag aaaaaattta aaaagataag tagaaaaatg aagtaactgt agaaaccata 1080

cttactcttt gatctcaaat gcccaaaaac tgaatgaaaa tgtgaattta ggccgaccag 1140

gtagtcttgt caataaacta aaagaaaaac aggaaaattg agaaatatgt tacaactata 1200

acaacacaaa acagcatagt tttgaaacac ttgcagttct taaatataaa agcttttatt 1260

agttaatttt ttaaaaggat ctcataggat tgacactgaa tcaggttggg aggtggaata 1320

agggtgatgg catattcttt ctgaattact tattataaca tttctagaat cattaggtca 1380

gtgctacttt gttgtcgtca atgtacaata aaggaatcac aaattgatct tagtgataat 1440

tttacagagg cagacattgc acataggtat gactgcaaaa atgggtggct aactctggga 1500

agatacttgt gttaaacttt atatgacatt taataaccct tcatcataag gcaatgtttt 1560

ttacaaaaag attgcacaaa atcatgttag tcatttactc tgcaaaaatg gcacattagt 1620

gggggttcca aaatccataa tga 1643

<210> 78
<211> 755
<212> DNA
<213> Homo sapien

<400> 78
cgaggtataa aaactacgtc actctaaaat gttacaaata ggtcatctac ttagtatgca 60

tagccttgat aaaaacattg gtcaagtcgg gatgtagtcg gccaccaact agaaatgtgt 120

taagattttt ttaagcagac ttgcttaata aggcaaggag tggggtcagg ttgttctagg 180

ggccagcaga agggtctaaa atacagggta gtgaaaagag attacgagac tagtgagttt 240

cctttaaatg cttaactagt cattattaag acagccacat ttcagtgggg ctgagccaaa 300

ctgctgagct tggaatagca tatgcttgga atctgaatat gaataaggcc caggtgccac 360

actttacacc acagatcctt tgctaaagag gcactatttt gtctaacagg caaggaccag 420

gctggcagtc aggaaggctg ggtttcggtg ctgatcttgt caccaactat gcactcttga 480

acaagtcact tcacttcact atcctaagcc tgttatctca tctgaacaaa taacaggggt 540

tagacttagc cttttacaat gacattttgt atatatctac tgagctctaa caattattac 600

aacatatcta tgtctgacag ataggatagt cctacatatt caggaaactc cacgtatagc 660

tctcctaaaa ctgattgttg cgtgttacca cacaacacaa caacatacaa acctgggcac 720

tggcaacacg accggtcaat tctcccaaca caacc 755

<210> 79
<211> 1002
<212> DNA
<213> Homo sapien

<400> 79

tatttcatct ttatagggaa tttgctccca aggtatattc ggcacgagaa aaaacctcat 60

atttaaaaac tacgtcactc taaaatgtta caaataggtc atcttcttag tatgcatagc 120

cttgataaaa acattggtca agtcgggatg tagtcggcca ccaactagaa atgtgttaag 180

atttttttaa gcagacttgc ttaataaggc aaggagtggg gtcaggttgt tctaggggcc 240

agcagaaggg tctaaaatac agggtagtga aaagagatta cgagactagt gagtttcctt 300

taaatgctta actagtcatt attaagacag ccacatttca gtggggctga gccaaactgc 360

tgagcttgga atagcatatg cttggaatct gaatatgaat aaggcccagg tgccacactt 420

tacaccacag atcctttgct aaagaggcac tatttgtcta acaggcaagg accaggctgg 480

cagtcaggaa ggctgggttt tggtgctgat cttgtcacca actatgcact cttgaacaag 540

tcacttcact tcactatcct aagcctgttt tctcatctga aaaataaagg ggttagactt 600

agccttttaa atgacatttt tgtatatttc tactggctat aaaattatta caaatatcta 660

tgtctgacgg taagatagtc taaatattca ggaaaactcc aagtatagct ctcctaaaaa 720

tgatatgttg cgtgttaaaa aaagaaaaaa aagaaaagaa gaagggggag gaaaaaataa 780

aatgaaaaaa acttcaaaaa tgcacggctg agttggtagc aaagaaggaa attctttgga 840

ggccaaaaag atctagaaag tttaaatcca atgtgcagga gctggcattg cctagctaat 900

ccctcatgat tgagaacctc agattataga cactcatggg gaccaagaga taaggcctgg 960

ggcctcaaaa aggccagagc cgaggtcgga tcaaagaatc cc 1002

<210> 80
<211> 374
<212> DNA
<213> Homo sapien

<400> 80
tcttttctaa aactttaatt tccactatgg ctcttttgaa accattttaa tcaagtcaca 60

tttcttagaa aaaattcact cagggttctg aaggaattag ttattttcta caagcaactc 120

tgtcatgagt gatagagttg tagctctctt agaagttttt ttcccctttc aaagagaatg 180

agaaatatgc agagatttcc ttactgactc actaaatgta aagattaaga ggacataata 240

aaatttggga ctacagtagc atataggttt tcagtttatt tactactaac tagctataac 300

ttagacaagt catttaacat gctgtgcttt agtttcatct ttgaaaccaa agagattcga 360

accagaaatc tctt 374

<210> 81
<211> 399
<212> DNA

<213> Homo sapien

<400> 81
atggggaatt ccattgacac agtcagatat ggcaaagaat cagatttagg ggatgttagt 60

gaagaacatg gtgaatggaa taaggaaagc tcaaataacg agcaggacaa tagtctgctt 120

gaacagtatt taacttcagt tcaacagctg gaagatgctg atgagaggac caattttgat 180

acagagacaa gagatagcaa acttcacatt gcttgtttcc cagtacagtt agatacattg 240

tctgacggtg cttctgtaga tgagagtcat ggcatatctc ctcctttgca aggtgaaatt 300

agccagacac aagagaattc taaattaaat gcagaagttc aagggcagca gccagaatgt 360

gattctacat ttcagctatt gcatgttggt gttactgtg 399

<210> 82
<211> 517
<212> DNA
<213> Homo sapien

<400> 82
gaaagtatat tgacgtaggt agtggagacg ccatgagttc ataatctgtc cagagtcgca 60

gtatgatgta tccggcaccc gacaggtcaa gaaagaacta cttgtttcta ggaagaacat 120

atgaagtgct taatttataa gcgggctgtc gaatattatc caatatagtt tcttctgaaa 180

agtgaaaggg gatcatctat tgttagatta gggggtctcg gaaacttttt gaaaattcga 240

atcagtggac caatgtacat gtgaaaacta aagagggcag gggttaaaat agggcttgaa 300

tttctcattc tgtatagacc agcaaacttc cctgtgcaag gcaagtttac atcacaaatc 360

caagaatgtt tgcatcctaa atgctagttt gcttcagccc ctagttaacc tcaggacttg 420

gtttgcatat aaaaggtaga cagctgatat gttttcatga ataaatattg tcagccagaa 480

aaggttggtg tcaggtaatg catatttttt taagctt 517

<210> 83
<211> 619
<212> DNA
<213> Homo sapien

<400> 83
acacaatgat acccattttt gcatgttaat gtattattaa atatcagtgg gaatagtctg 60

catgctattt cacatctcag gcacacttaa ggaagacctt gtgatgtgca tgttgctcat 120

ttaatctaga aaggatacca agattcattt agaacttctt tatgcacagt ttttttttga 180

gtatgttatg tcctgaggca ttaagggtat tactaaagca agcagcggga cttctcagag 240

aaattaaagg tttcatatca accacacgtt gtcaaaatct tcactttgaa taggattaaa 300

tgatgtttca tcagtattct tggcacacat gacattgttt ttaaaataac agttttatta 360

ctctgggctg tgacagtttc tcagactttc cttaatatca tacaattctc caatttaaac 420

tggtatagtc agttttacaa tattttaatt accctgtatt cattagcact ttcctcattt 480

tctactacct cctccccagc tgcccctacc ctaggcaatg ccaaatctac tttctgtcta 540

tatatttgcc tattcttgaa atttcaaata aatggaatcg tataatacaa acaaaaaaca 600

ggaaaaaaaa aaaaaaaag 619

<210> 84
<211> 646
<212> DNA
<213> Homo sapien

<400> 84
aatgatccat ataggcgaat ggtcatctaa atcatgctcg agcggcgcag tgtgatggat 60

cggcgccggg caggtaactc accccccagg atagagaagt gtttgttagg gagagaagag 120

ggagaggcag gagccggccc aagcccaggg tccctgcttg ggccccagaa agcacttaac 180

caggccccaa gccttcaagg gaaaccaagg cctcaaccag acaatcttga gggaaggaaa 240

agccagactt tgggtttgtt ttttgggggga attattggtt tttttttttt tatgtttctt 300

ttggaatttt gtttgttggc aaattctgtg tgatcttttt tcataaaaaa aaagacaaag 360

aatttacatt ggacaaaatt aaaaaaaaac aaaaaaacaa aacaaaacaa acaggcgtgg 420

gcggtctacc tcaggtggcc atatgccggt gtgtcccggt ggtggtgaaa catgtggtgt 480

tatctccggc ctcaacaaat tctcccccac aacaattccg tccaccgcac caagcccgat 540

ctaacaacag gacatcatat agcaacctat atacgagcac ctcaacagca ccaacgacag 600

ccaagcgaga cgaacgacca acagacacac cactcacaac caaagc 646

<210> 85
<211> 419
<212> DNA
<213> Homo sapien

<400> 85
cggccgccgg gcaggtactt tcgttgatac aggcgtggaa gaccttgagt tcccctgtgg 60

ctaccccatc atagttcctc ctaaggctat accagataag ccatacggag cagatgacca 120

gcaagaacct ttccagaatt attattctaa ctagaatctt agccaagaga atggaatcac 180

cacaaatgtt atcatgaaaa tcatctcaag taaatttcct attccattca taccgttaag 240

ttgaggctcg atgatatacg aaaactttaa ctgaattgac ttcataaagg cttaatggtc 300

ttcaaaatta tgctggttat atgaattctt aaattcaagc tcttttccaa ataataaatg 360

ataaaacaac attttaatta gtattttacg taaaaatata tattaaaaag taaatcaag 419

<210> 86
<211> 2133
<212> DNA
<213> Homo sapien

<400> 86
ggaagtacag gataatatta aagtcaaata gagtacagtt cttcagcatc ataaatcaaa 60

attcaattgc tacaaaaatc aaaacttgtc agactttttg ctttaataca aatagttgga 120

atttctgagc aatcaggttt atctttaaat atgttttttt ctgagctttt ttacttcaaa 180

aacgatgaga attatcaatt tttcagtact actgacttgt tccttgtgga aggagggaac 240

attaagtatt taaatcaatt tcttaagtct tcgagtatca aatttatttt gtttaatctt 300

tgatttaatg tttaacatgg gcacttttta tattctctta cctgagttag ttttgaattc 360

ctagaacatg tccattttaa cagtggttgt gatattattt agttaatact actgtctgga 420

ttattttaaa atcttggtac aatttgtata aaacaacata acacttgtta acttgccagt 480

cctctaggaa cttgtttcct ttccttactc tgaatagact agtggtagct gtccattatc 540

ttttacctta attacgattg tttgaaccac atttaaattc caaaatctat attattggtt 600

taaaagcttc aacttgacaa gatattatta acagtctaca tgaaatcctc aaattatata 660

tgaattttca aacattgata tcagctcctt gatttacttt ttaatatata tttttacgta 720

aaatactaat taaaatgttg ttttatcatt tattatttgg aaaagagctt gaatttaaga 780

attcatataa ccagcataat tttgaagacc attaagcctt tatgaagtca attcagttaa 840

agttttcgta tatcatcgag cctcaactta acggtatgaa tggaatagga aatttacttg 900

agatgatttt catgataaca tttgtggtga ttccattctc ttggctaaga ttctagttag 960

aataataatt ctggaaaggt tcttgctggt catctgctcc gtatggctta tctggtatag 1020

ccttaggagg aactatgatg gggtagccaa aggggaactc aaggtcttcc acccctgtat 1080

caacgaaagt actctaatgt ctgttttaca tactgggatt atttgtaaga tttcatttga 1140

aaggaaggtt ctttagacca agaaagaaaa ggaaaaaggt tgaaaccaat gacctgctcc 1200

aatctcttag aaactgaatc tcagagaagt taacttcaag gtaaaagcat ttgttagtgc 1260

tagaggtaag atgaaaattc aagttttttt attccttgtc ttcataataa tataattatt 1320

gtgatgtctt ttgtacaatt tgcataatac tatgtataca ttcacatgta gtatttaagt 1380

tacataagtg atgggtacta tgaaattact attgatcaag aatgactatt agattttaat 1440

taagattaca ctttatttct tgtaaaaggt gatttaaaat gcacattcct taccaatcta 1500

atttgaatca tgattagcct cagtttaatt atccttacaa aaatattttt gagtggttgg 1560

gatcagtttt aagttgagct cctagatttg ttgaatagga aaggatacta ataactgttc 1620

taggggaaat gattttgtaa tatttcacct tgaatttttg aactgaacct tataaactag 1680

tcttcagaat gactaagcag gttaaatgtt ttagcattta aatgtcaaat agagaaatca 1740

atctgacttt tggaaaaaag aaagatgttt aatttaaaat atgtaaagca aacttccaaa 1800

tttcttccat cagtaagagt aactaactgt ctgaatgtag ttattattat tgtgtcaagt 1860

taaatgattg tacatacttt cctttacaga tttggataag tgaagacagt aataacattg 1920

aagcagtgaa ccagtggaaa gagacagtaa taaatccaga aaaggttgtt atcaggtggc 1980

acaaattaaa tccatcttga agacttcaca cattaatttg gtgaagaact tgacattctt 2040

ttagaagact tatgatttca atttgctacc aatgagaaga ggcaaatcaa caaatttgtc 2100

aatttatggg ggctataatt atggtatata atg 2133

<210> 87
<211> 493
<212> DNA
<213> Homo sapien

<400> 87
gcggccgccg ggcaggtctt cgatctcccg gggtgctggg attacaggtg tgagccacag 60

cacctagcct taccttcaaa ttctaaacca agctatttaa atagccactg tttgattatt 120

tgaattaaca tggagcatct tctgggatat tgttcaggga aatatgagta gatcaaggta 180

ttttggggat gtaaaccctc atgtttgata aaataaatga tattttgagc tactgtttgc 240

tgggaacaga aagtaagaag ggaaaaggag cgaccataca ggaaagtaaa aataataaaa 300

gaaaatttag aaaactagag gaaaaggtat gaaaggataa atcctccatc ccatactgat 360

aatggccttt gagcatcact aagccccttt gcttctccca ttaagcaaag gatgatgact 420

gaggaggaac aaacaaaaat agacatcatt ataaaaaata cccaagactt ttagatgttt 480

ctctaacatt tgg 493

<210> 88
<211> 1412
<212> DNA
<213> Homo sapien

<400> 88
tgaattagcc atacaaaaaa aataaaaaat tactgttagt caccctacag tgcaaggtaa 60

cactagaatt tatctttcca tctagtaacc actgtttttt aaagagacag agtatctccc 120

tgttgcccca gctggagtgc agtggcacaa tcatagttca ccacaccctg gaactcctgg 180

gctaagggat cctccttagc ctcagcctcc caagtagcta ggtatacagg catgtgctac 240

catgcctggc taattaaaaa agattttttt agagatgagg tcttgctgtg ttgcccaggc 300

tggtctcaaa ctcctgggct caaacaatcc tcccaccttg gcctcccaaa gtgctgggat 360

tacaggtgtg agccacagca cctagcctta ccttcaaatt ctaaaccaag ctatttaaat 420

agccactgtt tgattatttg aattaacatg gagcatcttc tgggatattg ttcagggaaa 480

tatgagtaga tcaaggtatt ttggggatgt aaaccctcat gtttgataaa ataaatgata 540

ttttgagcta gtgtttgctg ggaacagaaa gtaagaaggg aaaaggagcg accatacagg 600

aaagtaaaaa taataaaaga aaatttagaa aactagagga aaaggtatga aaggataaat 660

cctccatccc atactgataa tggcctttga gcatcactaa gccccttttgc ttctcccatt 720

aagcaaagga tgatgactga ggaggaacaa acaaaaatag acatcattag aaaaaatacc 780

caagactttt agatgtttct ctaacatttt ggggtcattt tcagattacc agtgttcatt 840

tgctgaggta tattaacgga tatttgtact taatttgaaa aatagcagga tccaaaccag 900

aggtctgtat aagagcaggc ggcatgcgtg tctggagagc tgctgcctcc acaagtattc 960

tgacagcact gggctgctag tgagacctgg atggccaccc tccccatgtc atggccatgg 1020

gttttcggga accgtttcct ccttttactg catcacagtt gcaaactcgt ctatttattt 1080

ttctcttgat taacaactgc actctgacat tgcagcagtg ttgatgaaga caatttaact 1140

catgtttttg ttaacataat aattgtctgt cgtaactaaa atataagttt cttgaaagct 1200

ataatcaggt atagagaaaa tctttgttat gcacaatacc agggcaggta atatctgtaa 1260

tatgtattaa cagcaattca ctaaacattg aatgtctctg tatgctggca cctgtgctaa 1320

agatttgctg tataaagata aataggaaat tgcctcttct cccacgaaac tcaaaacatt 1380

tattgaatga ataaataata ggtgaattaa ta 1412

<210> 89
<211> 624
<212> DNA
<213> Homo sapien

<400> 89
ggtacttgag gtgtttctca ggttccagaa catccgtgtc atcttaccag atccttcaag 60

gattcagctt aaagatcagc tccaccagga agccttcctg gatttcccct cttagtttcc 120

aacaagaatc cggctcttcc gttctctgcc caccttggag tagcagtagc gttcagctgt 180

gagactctcc gtgtttttcc cgttacagtc gtttgttagc gtgcatcctc tttcgactga 240

attagttaga tgtgagaccc taggactctc ttgttttctt cgttacagtc tttgttgctg 300

catcctctct cactgaattg ttgaattgtg agaccctgtg agggtcggca ccctgtgata 360

ctggccagga aagggttgtt gcaaggggat catgggattg ttgaatgggt tttgatctgg 420

attttgatgt tggaaatcaa gttcccaaat gttttcaacc ttgggtaaag gaacatgtaa 480

tggtgttttt taacaaaaca aaaaattaaa aaaaaaaaaa aacaaacata aaaaacaacc 540

aacggctggg ggcacccggg ggcaaagggg gcccgggggg acattgtttt tcccggtaaa 600

atccccaaat tgggaaaaaa aagt 624

<210> 90
<211> 659
<212> DNA
<213> Homo sapien

<400> 90
accacgcctg tagcctctgt ctagagtagt tcacacatgg atgctgtctc tctggtactt 60

gggtgtttct caggttccag aacatccgtt catcttacca gtccttcaag gttcagctta 120

aagatcagct ccaccaggaa gccttcctgg atttcccctc ttagtttcca acagaatccg 180

tctcttccgt tctctgccca ccttgagtag cagtagcttc agctgtagac tctcctgttt 240

ttcccttaca gtctttgttg ctgcatcctc tttcactgaa ttgttgatgt gagaccctag 300

actctcctgt tttttcgtta cagtctttgt tgctgcatcc tctctcactg aattgttgaa 360

ttgtgagacc ctgtgagggt cggcaccctg tgatactggc caggaaaggg ttgttgcaag 420

gggatcatgg gattgttgaa tgggttttga tctggatttt gatgttggaa atcaagttcc 480

caaatgtttt caaccttggg taaaggaaca tgtaatggtg tttttttaaca aaacaaaaaa 540

ttaaaaaaaa aaaaaaacaa acataaaaaa caaccaacgg ctgggggcac ccgggggcaa 600

agggggcccg gggggacatt gtttttcccg gtaaaatccc caaattggga aaaaaaagt 659

<210> 91
<211> 556
<212> DNA
<213> Homo sapien

<400> 91
aattttcaac tggcccatac tttatagtga tggaaagcgc ataacactac ttgtaaatca 60

ttaaaatagg gtgataactg tgataatagt gtttcttgca ttctagaaaa ttattttatt 120

aactacattc aaaacccagc atttcacagg ttccatcatt agaaacagta tagttctagt 180

taacatgatt ggagagtttc aggggaaagg tttacatttt ctgaaactgt atttggtatg 240

tgactcaatg tggtatttca gtcttgttag tcacttacat gactgacgtt tgcaaggatt 300

tattgccaag taaaatttga tcagagtgca ctgagaatag ctacataagg ggaaatctct 360

caaaattcct tctgttcact ttaattcgga gcatatgttt caactcattt tcacacatct 420

gtcccacagt tgaagcatta acacacatct tcacgacaca atgaacacat acacattagc 480

aaacataagt ctcttaatgc aaaattacta gttgactaca atatagctac cttaaaagca 540

gagcttgcta taattc 556

<210> 92
<211> 635
<212> DNA
<213> Homo sapien

<400> 92
acaaaatata atgttataaa tgtattttaa aaaaagaaat acaaattcta tggtcttttg 60

cattttactg cctcaaagca gaattagcaa agctgatgaa gaatgaacat tttcccttgg 120

gcgggtggcc cttggtcact cccacaggca cgttaccggg ctccggcgtg tgctcccacc 180

aaccacggca aacaaaggcg tcctcctcac ttgaagtcct ggcctgtggt tgtttcatct 240

gtttttttgc tcagtgaaca aaacgttctg aaattagaac tcaccaaagt taaaagcagt 300

aaaacaacat atgctactta agacattttg aagcggaaag taaagctatg tgaatgccgt 360

ccttccttcc ttcctttttc tacagcttgg aaacctctga gaatttgctg gcgggtggca 420

gaggagggct tcgtctagct cttgaacgga caggaactgt ctggctagac agctctccag 480

accacgaaag cccaggaggt gccctcttcc acacaacaga ctaagcactg cacccacttt 540

ctttgatcca gaaagcatcc ctactgaccc tgtaacctac accctctctg tccaaagaac 600

agaggccgac cagagtagcc agcctggaga ggcac 635

<210> 93
<211> 8156
<212> DNA
<213> Homo sapien

<400> 93
cggggcgtgc gcgtcctcct ccccaggccc gccgcctccc tgccaagaat ctgagagagg 60

ccgagtggag ttcggtcctt ctctgaacag ttttagctga gagtaccagc atccaactgg 120

gagcgttgtc attgcatttc cacattccca ggaaagccca ggtgctggct gccagctgct 180

gcgcccccca tgtagaaggt gcacctcctg ggagcaggca cgtcttttgg ctcttctgac 240

catggagaga taggacggtc cctgcagccc gcgcgacaga aagctgtgcc gccaccaccg 300

gccgcgtccg tccttcggat ggatcgcaac agagaggccg agatggagct gaggcgaggc 360

cccagcccca ccagggccgg ccggggccac gaggtggatg gggacaaggc tacctgccac 420
acctgctgca tctgcggcaa gagcttcccc ttccagagct cgctttcgca gcacatgcgc 480
aagcacacgg gcgagaagcc ctacaagtgt ccctactgcg accaccgggc ttcccagaag 540
ggcaacctga agattcacat ccggagccac cgcacgggga ctctgattca gggacacgag 600
ccggaggcgg gcgaggcgcc gctgggtgag atgcgcgcct ccgagggcct ggacgcctgc 660
gccagcccca ccaagagcgc ctctgcctgc aaccggctgc tgaacggggc ctcgcaggcc 720
gacggcgcca gggtcctgaa cggggcctcg caggccgaca gcggcagagt cctgctgcgg 780
agcagcaaga agggggcaga ggggtccgca tgcgccccgg gggaggccaa ggcagcggtc 840
cagtgctcct tctgcaagag ccagttcgag cgtaagaagg acctggagct gcacgtgcac 900
caggcgcaca agccgttcaa gtgcaggctg tgcagctacg cgacgctgcg ggaggagtcg 960
ctgctgagcc acatcgagag ggaccacatc accgcgcagg ggcccggcag cggcgaggcc 1020
tgcgtggaga acggcaagcc cgagctgagc cccggggagt tcccgtgcga ggtgtgtggc 1080
caggccttca gccagacctg gttcctgaag gcgcacatga agaagcaccg ggctccttc 1140
gaccacggct gccacatctg cggccgtagg ttcaaggagc cctggttcct caagaaccac 1200
atgaaggcgc acggccccaa gacgggcagc aagaacaggc ccaagagtga gctggacccc 1260
atcgccacca tcaacaacgt ggtccaggag gaggtgatcg tcgccggcct gagcctctac 1320
gaggtctgcg ccaagtgcgg gaacctgttt acaaacctgg acagcttgaa cgcccacaat 1380
gccatccacc gcagagtcga ggccagccgc acgcgcgccc cggccgagga gggggcggag 1440
gggccctcgg acaccaagca gttctttctc cagtgcctga acctgaggcc gtcggcggcc 1500
ggcgactcgt gccctggcac gcaggccgga cggcgggtgg ctgagctgga cccggtcaac 1560
agctaccagg cctggcagct ggccacgcgg ggtaaggtgg ccgagccggc cgagtacctc 1620
aagtacgggg cctgggacga ggcgctggcc ggggacgtgg ccttcgacaa ggacaggcgc 1680
gagtacgtcc tggtgagcca ggagaagcgc aagcgtgagc aggatgcacc agccgcgcag 1740
gggcccccgc ggaagcgcgc gagcgggcct ggggaccccg cgcccgccgg ccacctcgat 1800
ccccgctcgg ccgcgcgccc caaccgcagg gccgcagcca ccaccggcca gggcaagtcc 1860
tccgagtgct tcgagtgcgg caagatcttc cgcacctatc atcagatggt gctgcactca 1920
cgcgtgcatc gccgcgcgcg ccgcgagagg gacagtgacg gggacagggc ggcgcgggcc 1980
cgctgcggat cactcagtga gggtgactcg gcctcccagc ccagcagccc tggctccgcc 2040
tgtgccgctg ctgactcccc gggctctggc ctggccgacg aggctgccga agacagtggt 2100
gaggagggcg cccctgaacc tgcaccaggg ggacagccgc gccgctgctg cttttccgaa 2160

gaggtgactt cgaccgagct ctccagtgga gaccagagtc acaagatggg agataacgcc 2220
tcggaaagag acaccggcga gtccaaggca gggatcgcag cttctgtgtc catacttgaa 2280
aacagtagca gagagacttc tagaaggcaa gagcagcaca gattttctat ggacttaaag 2340
atgccagcat ttcaccccaa gcaggaggtg cccgtccctg gtgatggtgt ggagttccct 2400
tccagtacgg gagcggaggg ccagacgggt caccctgcag aaaagctgtc cgatttgcac 2460
aacaaggaac actctggggg agggaagcgg gcgctggccc cagacctcat gccgctagat 2520
ttaagtgcga ggtcgacgcg ggatgacccc agcaataagg agacggcctc ctccctgcag 2580
gcggctttag tcgttcaccc gtgtccttac tgcagccaca agacctacta ccccgaggtc 2640
ctgtggatgc acaaacgcat ctggcaccgt gtcagctgca actccgtggc tcccccgtgg 2700
attcagccca atggttacaa aagcatcaga agcaatttgg ttttcctttc ccggagcgga 2760
cgcacgggcc ccccgcctgc cctcggtggc aaagaatgcc agcctttgct ccttgctcgg 2820
ttcacccgca ctcaggtgcc aggggggatg ccggggtcca aaagtggctc ttctcccctg 2880
ggagtggtca caaaagccgc tagcatgcct aagaataagg agagccattc cggaggtccc 2940
tgcgctctgt gggcgcccgg ccctgacggg tatcgacaga ccaaaccttg tcacggccag 3000
gagccacatg gcgcggccac acaggggccc ctggccaagc ccaggcagga ggctagctcc 3060
aaaccggtgc ctgccccggg tggcgggggc ttcagcagga gcgccacccc tacgcccacc 3120
gtcatcgccc gggctggcgc gcagccctcg gccaatagca agcctgtgga gaagtttggg 3180
gtccccccag cgggggctgg ctttgccccc acaaataagc acagtgcccc ggactccctg 3240
aaagccaaat tcagtgctca gcctcagggt ccacctcctg caaagggcga agggggcgct 3300
cctcctctac ctccccgcga gcccccctcg aaggcagccc aggagctgag gactctggcc 3360
acctgtgctg cggggtccag gggcgacgcg gccttgcagg cccagcccgg cgtggctggg 3420
gcgccccccc gtcctacact ccatcaaaca ggagccagtg gccgaggggc atgagaagcg 3480
cctggacatc ctcaacatct ttaagacgta cattccaaag gactttgcga ccctctacca 3540
gggatggggt gtcagcggcc ctgggttgga gcacagaggg acactccgga cgcaggcccg 3600
gccaggagag ttcgtctgca tcgagtgcgg aaagagcttc caccagcccg gccacctcag 3660
ggcccacatg cgggcacact cagtggtgtt tgagtccgat gggcctcggg gttctgaagt 3720
tcataccacc tccgcagacg cccccaaaca agggagagac cattctaaca caggtaccgt 3780
ccagacagtg cctctgagaa agggaaccta aaggcgtgtt tccgacgcac cccaggtccc 3840
cgtaacggcc attagcagta ccctcacgat gtcccagcag cctcccacct gtgacctggc 3900

cgctccatgg aagaacagcc ggggaactcc tgagcagaca cctcacatcc cgagccgctg 3960

cgctggagtg gaaactgaag gcagatgcct ctccttgtta aacgttcaga aataaatgaa 4020

gatgctatat tctagaaata catgtagata ctatatacgc atttacgtgc tcatcgtcca 4080

tagtcccata ttttcttata ataaacagta gtactggcag gcacagtagg ggcacaaggc 4140

atctgtctta ttcaagacaa gtttgagaca ctggaaaaaa agatacttgt tgtgtgtgtt 4200

ggacagagtg gcgaggctga gcactgtcac aggggcctcc catgttaaga gggactgtgg 4260

ggatgatgtc agaacaagac gtggtggatt tgaggttgat cgagtattaa tactactgcc 4320

tctccttgtc ttagtgggta tttaaaatag taaataagag agaggaagga ggtgacgttc 4380

aggtgctgtg ggaagcaggc ttggcggagg ggtatgatga tgagaccctc attgttcact 4440

ggctccatcg cactcctccc tggggccgtg tgcctgttcc attcttccca ccattcgaac 4500

tgagcgaatc tggcaaagga gacacgtctg tgggaatgcg tagattccgc ctcggaagag 4560

agctagcgca acactaagaa aagcaggctt cttgtttatt ctcaggacct tttgtaaca 4620

gggctacatt ctgcaaactg cttacaaagg aagactatac gtcttaacaa attatttagc 4680

cactgagtcc tcccgattcg gacctgtttt agtaatggca gaagaatccc tgagcaggtt 4740

caggtgccct agatgactag ggtgctgagc tctggcgcct tctgtcccca ctctttgcct 4800

ccccgcccct tccctgagcc accccagcaa gtgggtgtct tttctccctg ggcctggtga 4860

cctccacagg atgagtgact ttgttcataa agggtgggga tcaccagccc cttgggtggg 4920

ggacggcttc atatacctct tcctcagtaa tgcaaatgcg agtttttgtg gtgggggtta 4980

aggcccataa caaaggatct taaaccatgc agtgtacgca attgaaatgg tattccacag 5040

atataaatat tttcttttcc cattgccgtg acactatgtg tgatggtaat atttctgaga 5100

gtttcagatt tttgcacata tgattttatg cattatcaaa agttactgct gccttgaatg 5160

aaaatgttct gtgaaatttt ttgcaaaagc tttactaggt tttttttaa ttgtgaaatt 5220

ttgtaaaggc aggaaatgga ttaaaacgag catgctaaat atatttttca aaaaagcaat 5280

aattttacat gtacagaaat tatcctaacc tttaatactg gcgagagcaa cagtttactt 5340

aatacggtaa tggactagtg cagtttttgt agacagtggg cttctgatac aaagtcttgt 5400

ttaaacacag acacacacac acacaaacac acacacacac cctaaagtgt gggtttcctg 5460

ttctaatgat ttgttgaata ttattatatt attattatta ttattattat tattgttatt 5520

gttattagta atgtttggtt ctggattcta cttgttactg agtttaaatt acttgacggt 5580

tcaggttact ttgcaacact ttcaaacgat gcaatgtaac tggctagctt atatatatat 5640

atatatatat atatatatat attttttttt ttttttactt attttttttct gatattctta 5700

caccagatat gtacgaaaat gatctgtcct gttggtgtaa ttaggaatgt ccatgcagat 5760
acagttaaac aactgtaatt gactgttctg taaagttatt ttgggcaaag ttgcggagac 5820
acattcctct gtccacctaa gaaatcagaa gactcttctg ttgatttatg tttaatcatt 5880
tcagtagttt ccccacagtg atcatttctg cattttctgg cttttgtttt cttggctgaa 5940
agtgaatggt gactgttagg aatgtcaggg actagtgacc cagtcctgtt tctctgtgtt 6000
ttagttatta aaaagaaatt ctgtacccaa agtgacacga aagtgtagtc tacattttta 6060
ctgtttcaga agcggcatgg aaaagtgcag ttggcctttg gagctggaag tgtcttgctg 6120
gtgaggctcc atcctggagg ctctggtggg gagtgggctg gcgctggggc cctgccggcc 6180
gcgtgctgga tccttcctgg cttgcaggag agcaggcgtg gaggacagtc agcttgcggg 6240
gccgcgcagg gtgcacagag tgcaggagga aggttttcac ccagttaaac agactgggga 6300
gccccccaa gaacgccatc ccttgaggcc agctgtgggc aggcctggat gtgtggtccc 6360
ttccttccac ccatcgtcag tattgtgttg gtttgttaat ttgttgattt ggtcatagta 6420
tttaatatga tttgtgtttc cttatttatt tagccaccgt tttgattgcc tttttttttc 6480
cgaatggtaa tttctgcatg atacacttct gtacgttgtc ttctgactgt tacagacttt 6540
ctactacctc tcccgatctg ctgtttcctt gtttcttaac aggatttttt acagtgttgc 6600
gtctaatgta acttagacaa taaagggttt ggttgtctac actgcagctt ctcggtgtct 6660
ctcccctgct ttccgctcgc tgcttcccgc tctgcccctg ctggggcctg gctgcaccct 6720
ggcctgcctt cctatactct cctgtttccc gctcatatct cttcctcatt tttgcgttca 6780
aataacacac agctaatgag cttctaaaaa tcttttcagg ttgttcactt gtattcctta 6840
atttgaagaa tgaatattta aattctctca aaagtcagat attgaggatc ttctctggga 6900
aattggccac tgtacctgcc cacctttctg cctggttccc tggaaggtct tattgtcatc 6960
ttagacggac agatttcatt ctcagcacca tacagatttg gcttcaaagc caggtgaatt 7020
ttgcctttga ggctctgaaa agtattaagt gttttaagag gtcccccaat atttacttat 7080
ttattttta aaccaagaaa gacactggtt ccctgaaaag caggtgcttc aggaagtagc 7140
aattgggagt tgcatacagt tacttcgtca gagaaaggag cgcccagtat gacaggcccc 7200
cacccctgat ccggccactg tgcacaggtc gctgagggtg tgagaacacc tctgcagggg 7260
ctccggcaca tgtgggtttc atcgtctcac actccttcag gctgcagggg ttgagtgcag 7320
aaagggcaag cttcatctcc atggtgcctc tccaggctgg caactctggt cggcctctgt 7380
tctttggaca gagagggtgt aggttacagg gtcagtaggg atgctttctg gatcaaagaa 7440

agtgggtgca gtgcttagtc tgttgtgtgg aagagggcac ctcctgggct ttcgtggctg 7500

gagactgtct accaacattc cttccttcaa actagacaaa ccctcctctg ccaccccag 7560

caaattctca gaggtttcca agctgtagaa aaaggaagga aggaaggacg gcattcacat 7620

agctttactt tccgcttcaa aatgtcttaa gtagcatatg ttgttttact gcttttaact 7680

ttggtgagtt ctaatttcag aacgttttgt tcactgagca aaaaaacaga tgaaacaacc 7740

acaggccagg acttcaagtg aggaggacgc ctttgtttgc cgtggttggt gggagcacac 7800

gccggagccc ggtaacgtgc ctgtgggagt gaccaagggc cacccgccca agggaaaatg 7860

ttcattcttc atcagctttg ctaattctgc tttgaggcag taaaatgcaa aagaccatag 7920

aatttgtatt tcttttttta aaatacaatt tataacatta tattttgtac tctttatatt 7980

agaatttgta actagattga tgtatttaac tatttctgaa aaagtaattc aatgttttag 8040

ttgtgtgata aaaatattta gataaaacat attcattcta ttggaatttg aaataaataa 8100

aaacatcttg gagttctgaa aaaaaaaaaa aaaaaaagat ctttaattaa gcggca 8156

<210> 94
<211> 668
<212> DNA
<213> Homo sapien

<400> 94
tggtcgcggc cgaggtatcc cttagaaatt gacagcttct atgaaattta caatcaagaa 60

ggcataagaa caactgctgc agcttcagaa ctatgcagaa aataaaatgt caacagctgg 120

gagaaaaaat tctcagtgag caacagagcc agtgaaaaat atcaccagta gagcaggcac 180

ctgagacagg gaaatggcac ccactaagtg caggtctaca ggggctgacc ttgcagacca 240

tttatggatc ccaagttacc aaaaccctta tttgattatg aactgcatag tagataccaa 300

tacatcagag cattgtctca atgttaatat ctatctgtgc tgattgtatt gtggtcatgt 360

aacagaatga atgcccttgt tcttaagtga tacgtgggaa aatatttgag ggtgaagtg 420

tcatgatgtt tgctacttac tctcacatgc tttggcaaaa ataaaacatc tctctctctc 480

caggcaaaat atagtaatca gtaaaatatt aacaatctgt aaaaccacac cacaaccaaa 540

caaaaaaggt tgggggacaa ccaagggcaa aagggtgttc ccggggtgaa atttgttttc 600

gggccaaaat tcccccacat ctcccgcaca aagcgggagc aaaaaaaacc acaaaaaaca 660

cacataca 668

<210> 95
<211> 746
<212> DNA

<213> Homo sapien

<400> 95

```
gactaagaca cctttctaga cagagaggag gccgatggca gacattctca gataggtttg      60
tagctattga cctggctgca tcaaaggaga tgaaatccct tagaaattga cagcttctat     120
gaaatttaca atcaagaagg cataagaaca actgctgcag cttcagaact atgcagaaaa     180
taaaatgtca acagctggga gaaaaaattc tcagtgagca acagagccag tgaaaaatat     240
caccagtaga gcaggcacct gagacaggga aatggcaccc actaagtgca ggtctacagg     300
ggctgacctt gcagaccatt tatggatccc aagttaccaa aacccttatt tgattatgaa     360
ctgcatagta gataccaata catcagagca ttgtctcaat gttaatatct atctgtgctg     420
attgtattgt ggtcatgtaa cagaatgaat gcccttgttc ttaagtgata cgtgggaaaa     480
tatttgaggg gtgaagtgtc atgatgtttg ctacttactc tcacatgctt tggcaaaaat     540
aaaacatctc tctctctcca ggcaaaatat agtaatcagt aaaatattaa caatctgtaa     600
aaccacacca caaccaaaca aaaaaggttg ggggacaacc aagggcaaaa gggtgttccc     660
ggggtgaaat ttgttttcgg gccaaaattc ccccacatct cccgcacaaa gcgggagcaa     720
aaaaaaccac aaaaaacaca cataca                                          746
```

<210> 96
<211> 978
<212> DNA
<213> Homo sapien

<400> 96

```
cggccgaggt accctgtgcc aactagggga ggagaactgt tgcctgctaa gttggtggca      60
ggaaccagct tcctgaagga tataaattcg ctggaagaaa gaccttggtt tatgttccag     120
tgctgttttc ccatctctag agcagtggct ggcagacatg tggctactca agaatggttc     180
ccagatgaat gaatgcagga aacagttccc tcctccaatg agattaacag ctgatccatg     240
cttataatga ctgaactctg taaagagggg aaccctctgc caatggggga tcaaaatggt     300
taatgagagc cctgctgtgg agagaggtag gacagcaagc acagaagcac ctgaccccat     360
gcagaggacg ggaggcagaa gcaggggcca gactggaggg agtgatctcc atatgcccat     420
atagcatcca tctgtcttgc caagcaccct tctggggtcc tctaccttta ccagagcata     480
gtctcttgtg cagatcaatg aaacgaagcg aagctggata gactatggcg agagcacttc     540
cctctctgcc tcccccaaga tgaatggcta cttgtggaga ggagctgtgg catataaaca     600
ctttcatcct tacccaaggt cccacatctc cctcaaatgg cataacaggc agacgagcgg     660
ccaaccaact ccctctgatt ctcatacaca gggtgtggcc ttccattggc atctttgagt     720
```

ggccccaagt cgtacgcaga ttgtggactc agtacacagc ttgcgcaaac ctggacaatg 780

tggccatggc ccatcataca cctctacacc acctcatagc gacgttgaat atgagatcca 840

cccgtagtgc ccagctcata acatcctctc cattaagatt gaccacaggc aacttaccat 900

tgactaggac caagtccccc aaacaccaaa attgagaaca gagcaacatg gtgccaaaca 960

tatcacagag aaatcaac 978

<210> 97
<211> 787
<212> DNA
<213> Homo sapien

<400> 97
acctggcaca aagcaaacaa taaatattat tgttattgtt gttataattg taaaatgaat 60

gacttcaaaa acatagtccc agtttggagg gattttgtga tgcagaatat ctaagtcata 120

gaaatagaag acaggtggaa taagtatatg ttcagagttt ttagatgtgt tgagtagaga 180

cggtaataat ggaagcatta aatacaaatg aaaatcacac cagatatccc tgaaattcaa 240

gcaaagaaag ttcatcatgt attcttgggc agcaagagaa aggactaggg ttatggcaat 300

gtgtggaaaa gttgaggctt gctaagggtt gagatctgtt ggtagccctg gatcacatgg 360

ggtcagcacc aggcagtgcc tctgaaagcg gagagaggtc ctggacttcc cttgtgtata 420

acagttccta gtgtccaaca atgaggaaac ggtgaagcat ggttacaaaa ctgtgacaaa 480

acatatttac atctagcact gttaccactc accatgccaa acattggctg cacacgtgca 540

gcccttattt gtaattaaca tcaaaagact agatctgaag ccttccataa atgagagacc 600

attcatatgg cattcctgga acaaaacact gcacaggtac caaggctctc cactccctga 660

cgggttggtg ctgaacagtc agggattgtc ttgactagac ttctgatgct tctgcatctt 720

ctttcctctt cccggaattc caaataacca attcatacca ttgtatttat gcttcgggta 780

acctagt 787

<210> 98
<211> 3670
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (3416)..(3416)
<223> a, c, g or t

<400> 98

```
agcggacagc cgctccctcg ctctgctggg gcctccggac gcgcttccca cgcgggtctc      60
tggaacactc ggtccgaacg cacgcctgct tgcactcaca ctgcggttca cacccggagg     120
cgctctcgca ctcacactgc cgctcacgcg cgctcacact cccccacgcg cgctccgctc     180
cggctccagc cccgcgccca gcgaaggcgc aggcactgct gccgagagcg ccgaggggcc     240
ccgcggcctt cccatggcgg acctgagctt catcgaagat accgtcgcct tccccgagaa     300
ggaagaggat gaggaggaag aagaggaggg tgtggagtgg ggctacgagg aaggtgttga     360
gtggggtctg gtgtttcctg atgctaatgg ggaataccag tctcctatta acctaaactc     420
aagagaggct aggtatgacc cctcgctgtt ggatgtccgc ctctccccaa attatgtggt     480
gtgccgagac tgtgaagtca ccaatgatgg acataccatt caggttatcc tgaagtcaaa     540
atcagttctt tcgggaggac cattgcctca agggcatgaa tttgaactgt acgaagtgag     600
atttcactgg ggaagagaaa accagcgtgg ttctgagcac acggttaatt tcaaagcttt     660
tcccatggag ctccatctga tccactggaa ctccactctg tttggcagca ttgatgaggc     720
tgtggggaag ccgcacggaa tcgccatcat tgctctgttt gttcagatag gaaaggaaca     780
tgttggcttg aaggctgtga ctgaaatcct ccaagatatt cagtataagg ggaagtccaa     840
aacaatacct tgctttaatc ctaacacttt attaccagac cctctgctgc gggattactg     900
ggtgtatgaa ggctctctca ccatcccacc ttgcagtgaa ggtgtcacct ggatattatt     960
ccgatacccт ttaactatat cccagctaca gatagaagaa tttcgaaggc tgaggacaca    1020
tgttaagggg gcagaacttg tggaaggctg tgatgggatt ttgggagaca actttcggcc    1080
cactcagcct cttagtgaca gagtcattag agctgcattt cagtagccaa agaggacagg    1140
aacaagtctg tcttcatgag ggaggaagac aatggtccta taatgccctt ggataagaaa    1200
aggaaacttt tgagctgcac cttcagttta tcctcaaagc ctgcgttgtt tgtcttcatc    1260
taatccagct ttgatggaca tctgtgatgg ttgcctgtac acttgctgaa atgaaatatt    1320
agaaatggct gtatattcca aagaaaccct atattatata tccacattac tgctgctagg    1380
attcatagtt gcacatactg tttattgctt atgtgtagaa ggaatgaaac tagtttccag    1440
agttgttatt aatatgaata tatatcatgt gttaatattg agaaaggaaa aatacattcc    1500
cggtgttagt agttcttcat ttcctgtctc caacagaaaa ttcactcatt ttagaactag    1560
tgtaattctt gataataaaa taagagtttt gattaagaac agcatagagc ttcaaaatgc    1620
aaagtgaatg attagtaaaa ttatgtctca ttttattttt tcagcaccca taccacaatt    1680
aatattaggc tggattgcca tgggaaacat tttttggcat taatgcagca acataatact    1740
cactttaggt attactacat agttgaagga tttaactgaa tgtatggatc aaatttattt    1800
```

atttgacata ttcgaagctg tggtttaata ggaatttgag aaaggtgtaa gaaataggat 1860
aaaaagaagg tcagcaccat gtaccaggaa tagctttact ttccatacat agaaatataa 1920
atttagtggt atcctatatt actttagtgt cgtacgcttt gtaagactta aatattttat 1980
tctattgatt ccactacttt ggtatgttaa gacatttctt taaagatgac caacaatatc 2040
cttattttag gtgccactag cagatgtaag cgtatactta gttgccgtta gatgtgacag 2100
aatgagataa tttatgtaaa gcagtagagt acctggcaca aagcaaacaa taaatattat 2160
tgttattgtt gttataattg taaaatgaat gacttcaaaa acatagtccc agtttggagg 2220
gatttgtgat gcagaatatc taagtcatag aaatagaaga caggtggaat aagtatatgt 2280
tcagagtttt tagatgtgtt gagtagagac ggtaataatg gaagcattaa atacaaatga 2340
aaatcacacc agatatccct gaaattcaag caaagaaagt tcatcatgta ttcttgggca 2400
gcaagagaaa ggactagggt tatggcaatg tgtggaaaag ttgaggcttg ctaagggttg 2460
agatctgttg gtagccctgg atcacatggg gtcagcacca ggcagtgcct ctgaaagcgg 2520
agagaggtcc tggacttccc ttgtgtataa cagttcctag tgtccaacaa tgaggaaacg 2580
gtgaagcatg gttacaaaac tgtgacaaaa atatttacat ctagcactgt taccactcac 2640
atgccaaaca ttggctgcac acgtgcagcc ttatttgtaa ttaacatcaa aagactagat 2700
ctgaagcctt ccataaatga gaggccattc atatggcatt cctggaacaa aacactgcac 2760
aggtaccagc ctctccactc ctgaccgggt tggtgctgaa cagtcaggga ttgttcttga 2820
actagacttc tgatgcttct tgcaatcttc tttcatcttt ccctgaaata cacaaaataa 2880
acaaatacaa taacaaatag taattaaatg actttcagga taacatctag ttgttcagac 2940
ttcacccttc acaggtgtgt gtgtatgtgt gtttatgtct gtatattgaa gcaatttgaa 3000
tttatttact gtatattttc tgagtaaaag actgaaatga actacttggt tcagatcatg 3060
gtgtccattg gtgacattgt ttggaggcat aatattcttt atatggaaaa tcctttaatt 3120
ccacagttag ttacctcaga ttcagaatat gaatactgtt tataatacgc ttttgtagga 3180
atgaattcga aaggtagttg tcagtaaaca aaagcacaac aaactaatct cagagtctgc 3240
cctgatggct gtgataggga cagaaagcta aaccctactg ctgacgcgcc ccgcacattg 3300
ggcgcagaat ttcccaagaa aacggggcaa atcaccgcca cggtcctaac tctgaactct 3360
atacgggcca tctcgcctaa accactacaa ggcacgcacg ggaaaggact ctccgntcgc 3420
gactcgcaag cctacggccc ccgaacgaca ggcgcaccac gacaccaccg gcgcgtctac 3480
gagacatgat cagcgtcaag ggcacctgaa aaaacgatgc cccaactagt gcggcccgca 3540

accaggcaga cactaagctt gatagcacag cgactgcacc aagagctaat cacgcacaca 3600

accaaagaca gaaactaccc actctatcac tacacggacg acactagaaa caacctgcaa 3660

ttgttactgc 3670

<210> 99
<211> 938
<212> DNA
<213> Homo sapien

<400> 99
ccaccccga cgacgacata ttaggggaac gggccactag atggctggtc gagcggcgca 60

gtgtgatgga tgcccgggca ggtacataat gttcagacct cctccatcct tttaaatgcc 120

tgctgcagta aataactagt ttgagtagaa ctagatcctg tctatctatt tggcacatgt 180

tctgctgcct ggggagtaag caagctaaag ggatgagaaa gaccacctcc ccctaccctg 240

gaaattgcac tgcaaggcag ggcgagaatg gggtagctgg cagacctggc ctccttgttc 300

ccagtcttag ttatttcttg cagagattca gtattcagta aagaatagca ttcaattagt 360

caaaaaatat atatctaact tcttcctttc ccttcccatg aatcattgca cgtcattccc 420

taagctttct tctctttcca cctcatggcc tgctcagtct tcccatccct accaatcaca 480

gactctcagc ctatagacgc agtcacagta tctcaactca tccgcctctg cttcacacta 540

cattaacaat acctcctcac tcacatacta cataactcca gctctagtct tccaaaattc 600

acctttcatg atgccactca gcatctcaaa tacctttcat gggctctctg ctgccaaagg 660

ataacaggtc aaagtcatta gcctcaacag tgggcttcaa ccagccttgg gacctcagcc 720

catttatcca tcacagaggc tggtaactag tctcactgct caggctgtga gtgttcctga 780

tccttgtgac attctgtgct gtgctttaca tggaacaggt ctttcctctc tctggcccat 840

tcgaatcctc taatcaagcc catctgattc tgtacagaac acattttcaa gttcaattcc 900

ctggatgcgg ttgcgcgaaa agttgcttaa tgactggg 938

<210> 100
<211> 376
<212> DNA
<213> Homo sapien

<400> 100
tactcttggt tttcttcctc caagactact ccttactcat atcagcaaat agcagctctt 60

ttcaagtgct cagtgtaaaa acctacaatt aatccttgat ttctctttca gtcagcctat 120

actaaatcaa tttcatttaa aatatctcgg ctactactct gcatctccac tgctaccatc 180

ggcctctcca gtcacattct ccaagagcac tctatctcat ttaaaagaca aaatctctgc 240

agtggcctgt gatgctcctt aatggcctac ataatccagc cctcaagcac ctccgtgatc 300

tctgtaaaac tttcccttgg tcactgtgct tcagccacat taaccagctt gcatatttct 360

cacattcacc aagctt 376

<210> 101
<211> 3661
<212> DNA
<213> Homo sapien

<400> 101
ggacacaact caacccagta acagttagtc aatggctgtg gcaggctaaa tgtggctccc 60

aaatatgtcc atatcctaat ccctacagcc tgtgaatatt accttatata gccaagagga 120

ttttgcagat gtgattctga gattgagaga ttatgccaga ttatccaggt aggccccaaa 180

tgtaatcacc acagtcctta taggagaggc aagaaagtca agtgtagaag gaggcgatag 240

aaggagagag ggatttgaag attaataggc tgcttgcttt gaagacagag ggaagggacc 300

atcaaccaga aataaacctc tagaagctgg aaaaggcatg gaaatagacc ctcccttaag 360

gtctctggag ggagtgcagc tttgatttct accgagtaaa attgattttg tacttcagac 420

ctccaaaact gtaagagaat gactgttgtt ttaaaaccat tgagtttgta gtaatttgtt 480

gcagcagcca caagaaactg gtacaacatc tatatagaat tttttcagat aattgggagg 540

aaatttgaat atggatggca tattaatatt actgaatcag cattaaattt gttaggtgta 600

ataatgtgat tgtagctatt taggagaata tcctattttt aagagacatg ccaccatatt 660

tagggagaag tgccaacata tttgcagttt attttcaaat ggttcagagg ctgtctgtgt 720

acatgagaag acaaagataa ggcaaatgca gcaaaattgt aataattggt gaatccaggt 780

gaagggacta tggctggtct ttgtactttt ttttccaact tttctgtagg tttaaaattt 840

tcaaaataaa aaatgggaaa tactttaaaa attgtaatca aagacattag tacagaaact 900

ttcataatgt attttatttt tacagtaaaa ttaatttatg taaattgata gaattttact 960

aatttcactc ccaagttaca ttaaaaggct tacatatgtt tgataatagc atatgtaaac 1020

tagaactctg aatgatatcc attggtcata atacgtacta tgtagcggta atggtgactt 1080

ttgtgattgc acaagtctag agatgcccca aatgacattg acttagacat ctggttattc 1140

taaggctgaa actgaagttg aatagaaggt tttagtcaaa tactgagatg aaaactgagg 1200

cagtcctggc gggggggagt gagtgtgtgt gtatatatac acacatagac atcatgcttc 1260

taaacattta cagaaagaaa gggtagatta tctacaaaaa aataagaatc agactgatat 1320

gagatcttac aaacctaacc cccttctctt tcctaaactc cagattctca tatttctgac 1380

ttcctatttg atatttacac ttcgatattt accaggagtc ttcaacattt tgttcaaaac 1440

agtactcttg gttttcttcc tccaagacta ctccttactc atatcagcaa atagcagctc 1500

ttttcaagtg ctcagtgtaa aaacctacaa ttaatccttg atttctcttt cagtcagcct 1560

atactaaatc aatttcattt aaaatatctc ggctactact ctgcatctcc actgctacca 1620

tcggcctctc cagtcacatt ctccaagagc actctatctc atttaaaaga caaaatctct 1680

gcagtggcct gtgatgctcc ttaatggcct acataatcca gccctcaagc acctccgtga 1740

tctctgtaaa actttccctt ggtcactgtg cttcagccac attaaccagc ttgcatattt 1800

ctcacattca ccaagcttgt tcctgccttg gggcctttgt acttaccatg ttctgttctg 1860

agaatactct gcctcaagat atcctacaac tatcttactg tattcagctc tctgctcaag 1920

tattaactga tgaaacctgt catccctact ccactccatg ttctgcttta cttaacagca 1980

attgcacata tggccccctg aataatatac atttagtcac ttattttac ttatctgcta 2040

attaaaatgt agactttttc tattctgttt actgctgtat tcccagcatg ttttatccga 2100

atgtgcagtg gtttcttttc ttctccctta tcgtgggaag tgatgtgcac aaatacacat 2160

aatggagcct gaatgtcata ttgctttcat acctgtgtga attttggtaa gaaaggaaaa 2220

gtagcgattg acaggtaata taattacatt aagtcactct catagttagc tgtttattgc 2280

tttcctgctc ttattctcag tccccaggac caaatgttga ccactacctt cccccacata 2340

taattaggtt atttaccgaa cgccatgcag gtggctgtta aaaggaagat atatacttac 2400

cttataaact caacttttcc ctgttgtctt tctgtctcac ccctacctcc atgctttaaa 2460

ttaacttttc aggcttaggc cttatctctc agtagagcca tataaggtat gtgtaaaagc 2520

aggaaaatgt ttcctgggga tgaagctttg aaaagctttt tttttttttc ttttggcaat 2580

aaaataaggt agattcagca caatacctaa taactaaaaa atctgttttt aattgggtgg 2640

ggcagacagc aagtgtgtca tcctggaaga tactatttgg gattttatgt aggtacataa 2700

gagaaaaaag tgaacaaaag caaggggcta ccaggacgcc gcagtatgct taacatgtat 2760

tttctaagtt tgtattatgc ctttatcttg gtacttttat cttctgttct cacttgatct 2820

ttttgaaatg tattttaaat cctaataaaa atatataaag tctggaatta ataaaggatt 2880

aaatgaaact tttgtatatc tcactgaaat tctcagaaaa aaggggggtg tggggagggg 2940

gaattgcctg ggtagtgag tgaaaattgt gaccaggttc ttactaagga atatggcaac 3000

tgcataatca aatgtcagtg gttaccaaac ttatgaatca cctggtgttg tgtcatagat 3060

tgtctatcct tgcctctcgc cccagtgat ttagatcagt ggaactatgt ggggtttaag 3120

aaatatacaa tatatatttg tatatatttg tgtgtctcga aagcttcagg gttaaataag 3180

ttttaactgt ttaggaaaca ctattgtttt aggtatccag tctcaaagac gaaggccttt 3240

aaaacttact taatttttca ttacatttct tgcccagaaa attgtaaaat acccaacgat 3300

aacaatgggg aattgtctat cagcacttga ctaaaaagct ttactatcca tgacagcagc 3360

ctttgcatta ctcaattctg atggcattta acgtcttgaa acccagaaat aaatacctat 3420

agactcacag tacctgaaag gaataccaaa ttgagacaag agagctatat aaaccaaaaa 3480

ttgcttcaac cacagaatgg aggtctacag gtgcggaagg aaagtttata tggtgaggct 3540

tggtcgcaaa actattagga atattttcag gttactacac aattttgcga gctcaatatg 3600

cagttaacac tttttccctc gaatctcctg agcagattta cattgaccgg acccgtagca 3660

t 3661

<210> 102
<211> 698
<212> DNA
<213> Homo sapien

<400> 102
acatttccat ttccaccggc ttggagcaga gctgtcgagg agtgctattc taggatcctg 60

atgatgacca caagggcagt ttgtattcag ctgtccctgg gaacacttcc ctgaaagcgc 120

tcagggacat tttcatcagg cacagtgctc caggctacgg cactctgtat tgttccctgg 180

tggctttagg gggctgggca tcgtagctga aataggacaa cagggagatg gctgagtgtg 240

tttcccaact gccagatgac aacaggtcta tcagcataaa gtcatcatat aacttagaag 300

aaaccttacc ctcggtgaaa tctcccagca gatcagcaac gaaatggact aagcaacttc 360

ggtagaaaca catggggcta ggatataaac agttcatagg aaaggacacc tgatatcatt 420

aatgattagg gagagaaatt gggtagctaa cagcaggggt gagagagaaa ctttatagta 480

ttttcctctg tagcttttga attttaagac atatgaatgg attttttttt taattgtaat 540

taaagtataa tttttttaaa agagaaattt ggagtcattt aacttgtaag acaaaggcta 600

tcttgtaata agaatactgt tcttcctatt tgctctagat tttaagtttg gatgggctac 660

atggtttctt agggcagaac cactcttata gactattt 698

<210> 103
<211> 1217
<212> DNA
<213> Homo sapien

<400> 103
acatttccat ttccaccggc ttggagcaga gctgtcgagg agtgctattc taggatcctg 60

atgatgacca caagggcagt ttgtattcag ctgtccctgg gaacacttcc ctgaaagcgc 120

tcagggacat tttcatcagg cacagtgctc caggctacgg cactctgtat tgttccctgg 180

tggctttagg gggctgggca tcgtagctga aataggacaa cagggagatg gtgagtgtgt 240

ttcccaactg cagatgacaa caggtctata agcataaagt catcatataa cttaaagaaa 300

ccttaccctc ggtgaaatct cccagcagat cagcaagaaa tagactaaca attcggtaga 360

aaaatggggc taggatataa acagttcata ggaaaggaca cctgatatca ttaatgatta 420

gggagagaaa ttgggtagct aacagcaggg gtgagagaga aactttatag tattttcctc 480

tgtagctttt gaattttaag acatatgaat ggattttttt tttaattgta attaaagtat 540

aattttttta aaagagaaat ttgggagtca tttaacttgt aagacaaagg ctatcttgta 600

ataagaatac tgttcttcct atttgctcta gattttaagt ttggatgggc atacatgggt 660

tttcttaggg cagaacccac tctactagac ctatttaacc ccatgacaga gcctagaagg 720

aacaggtgta atagaagatg gcatttatgg caagaaggtt gatcaagttc tccattagaa 780

tttgaaccag atctaatgcc ttttcttccc ttgtttaaga acggcccggg atgttggact 840

tcacgggcaa ggccaagtgg gatgcctgga atgagctgaa agggacttcc aaggaagatg 900

ccatgaaagc ttacatcaac aaagtagaag agctaaagaa aaaatacggg atatgagaga 960

ctggatttgg ttactgtgcc atgtgtttat cctaaactga gacaatgcct tgtttttttc 1020

taataccgtg gatggtggga attcgggaaa ataaccagtt aaaccagcta ctcaaggctg 1080

ctcaccatac ggctctaaca gattaggggc taaaacgatt actgactttc cttgagtagt 1140

ttttatctga aatcaattaa aagtgtattt gttactttaa aaaaaaaaaa aaaaaaaaag 1200

atctttaatt aagcggt 1217

<210> 104
<211> 193
<212> DNA
<213> Homo sapien

<400> 104
ccgggcaggt acaatatgga tttcaaaata acgttcactg gtaatccttc ctgatgccaa 60

ttttaaaatg aagaccgtct aaatttttct gaccagttat tagttgccct gcctctcgga 120

aatgtgttta aacttttctt tcaattattt gatacctttt gcccaagaga ttactatctc 180

tctctttttt ttt 193

<210> 105
<211> 542

<212> DNA
<213> Homo sapien

<400> 105
ggccgcactt tttttttttt ttttttagtt atatatttaa tgaatcattt ttattgcaaa 60

gggtaaatta catgaaattg acaaaattta gtccatgtaa tatctatcaa aatacataca 120

tgtaagtgtg tgtatattta tatatgtata cagtacagtt ttcacaaaaa gcttcaacat 180

tcctaagaaa cacagacata gtcattctgg tacaatatgg atttaaaata agttcatggt 240

aatccttcct gatgccaatt ttaaaatgaa gaccgtctaa atttttctga ccagttatta 300

gttgccctgc ctctcggaaa tgtgtttaaa cttttctttc aattatttga taccttttgc 360

ccaagagatt actatctctc tctttttttt ttttctttta agacagagtg ttgctctgtc 420

actcaggttg gagtgcagtg gcacaattcc tgatcactgc aacctctgcc tcccaggctc 480

aaacgatcct cccacctcag cctccccagt agctgggacc acaggcacat accaccaagc 540

tt 542

<210> 106
<211> 715
<212> DNA
<213> Homo sapien

<400> 106
ccgcccgggc aggtcctaaa tagaattcaa gattagacta aatgattttc agcagagcac 60

attcaaggtt ttacattcta tgattgaaaa aaattttttg aaaacttttt atttcattct 120

ttcctgtagg attttgctac aaataacttt gggaatgaat aaagtggaat ggtaactttc 180

cagtggttca gaattgaatt agacttcttg tgactgtgat gtttggtttc cattgaaata 240

tatgaagtga gatgtcatat cctgaatata gtttgtcttc cccaattact tgatagcatg 300

tctgtcagcc agtaaagatt aagaacagag tttctctaaa ttcctccgat tattccacta 360

aggcacatta aaatacttaa ttttgggaaa ccagacatca cagatttctc catgaagtcc 420

taaatcttct ttaaagtcag aataggtatc ttagttactg acagtattca ggtttttttc 480

tcccttggtg atatgtcatt ccatcagtga aaaaatattt tctcccaagg gatatagaaa 540

ggtattctgg taatacatta tcatcaatcc tttaacagta acagtctggc acttatcaca 600

aaaacgacca tttcttataa ccagaaagat atcttagatg tcttcacata tatttactat 660

gctgtagata aagatgcccg ggttatgggc tccatttcat ggcctgggtt acgtg 715

<210> 107
<211> 1716
<212> DNA

<213> Homo sapien

<220>
<221> misc_feature
<222> (1594)..(1594)
<223> a, c, g or t

<400> 107
agactgcaat ttctgactaa agcttttaat gccaggttaa acaggagaaa ctttttccac 60

tagaagaaaa tccttgctat ctattttttc caatagaaga aaatcctgct atttatttta 120

tttgatgaat aaacaaattt attgcagtag cttaaaaaaa tttttttttt aaacagtctc 180

actctgtcgc ccaggctgga gtgaagcaat gtgatctcag ctcactgcaa cctccacctc 240

ccgagtagct gggattacag acatgcacca ccaccctcag ctaatttttg tatttttagt 300

ggagacgggg tttcgccatg ttggccaggc tggtctttaa ctcctggcct tacgtgatcc 360

gcccccccctt ggccttccaa agtgctggga ttacaggtgt gagccactgc acctggcctg 420

tagtagctta aaattttcct tgagaaaatt cctgacttta aaaataaccc ttatataagt 480

acaagtgatt gtgacaaatg acgtaaaaat ggcattcatg atgtctgaaa caagcctaaa 540

tagaattcaa gattagacta aatgattttc acaaagcaca ttcaaggttt tacattctat 600

gattgaaaaa aatttttttga aaacttttta tttcattctt tcctgtagga ttttgctaca 660

aataactttg ggaatgaata aagtggaatg gtaactttcc agtggttcag aattgaatta 720

gacttcttgt gactgtgatg tttggtttcc attgaaatat atgaagtgag atgtcatatc 780

ctgaatatag tttgtcttcc ccaattactt gatagcatgt ctgtcagcca gtaaagatta 840

agaacagagt ttctctaaat tcctccgatt attccactaa ggcacattaa aatacttaat 900

tttgggaaac cagacatcac agatttctcc atgaagtcct aaatcttctt taaagtcaga 960

ataggtatct tagttactga cagtattcag gtttttttct cccttggtga tatgtcattc 1020

catcagtgaa aaaatatttt ctcccaggga taagaaaggt attctggtaa tacattatca 1080

tcaatcctta aacagtaaca gtcttggcac ttatcacaaa accgacccat ttcttataac 1140

cagaaagatt atcttagact gtccttcaca ttatacttta cctactgcct tgtaagaata 1200

agagttgctc actgtgttta cttgctgtcc tccatattct ccattgcacc attggtgtat 1260

aacgttaaga gtttcattga atattatttt aagtattaca aaaggcagct tgcttcttaa 1320

tctatgcatc tttggggttt ttgaagaaat ttaattcttt gatgtaaaaa ggaactgtta 1380

aaaaagttgg aagctctgca cctgtgtata tatatatttt agcaataaag cagcatgggc 1440

tgagaatgca ctgaaaaaaa aaaatgctag tgacttcagc aagtcataat cttcctgcgg 1500

gtggagggtc tcactgcgat gtggatggcc gctggggctg accagggtgg tggtggcaga 1560

aggctggggc ggctgtggca gtttcttaaa atangacaac aatgacattt gccacattga 1620

tagacttttc ttttcacaaa agaagtctct gtagcacgtg gttgctgttg gtgcacttta 1680

cccacagtgg aacttctttc aaatagtctc aatcct 1716

<210> 108
<211> 666
<212> DNA
<213> Homo sapien

<400> 108
tcgcggccga ggtacttaat aatgactgaa tttcatgttc ctacagtcat acatattcat 60

tagaagtttt atgttgttgg tctgatctga ttcttctttg tttgtgggtg gaacggcact 120

gagagaagta tagtttttta aacttgaaca tgttcagtag ttacattgcc ttagaaaacc 180

cagacacata gcagtggaaa tgaaagaaat ggcatcagaa gtgacttaat ttagcaattg 240

tgattcctct tgtaaaacaa aacaaaaaaa caatgccata tttttggag aaaagttggc 300

aatatagggg tttcgttgtc tgtttcacaa gaagactcat ttgttctttt gggggaacca 360

gtgccttaca gattttgtat atactgtaat tattcaggac tagggaacaa acaattgtat 420

tgtatttgtt acagattgta tatggctttg ttttaacatt cccctaaata aaatggcttc 480

attctcccct tggaaaaaaa catgactgtt atgttataaa acaaaaaaaa aaaaaaaaaa 540

aaaaaaggtg ggggtaccgg ggcaaaacgt gtcccggggg gaatggtttc ccggcccaca 600

aatcccccac attgcgagaa aaccgtgcga acaaaaaaaa aaaaaaaacg aaaaaaaaaa 660

acaggg 666

<210> 109
<211> 1983
<212> DNA
<213> Homo sapien

<400> 109
gaatttcgta atccttgaaa ttgaaaaaaa aaaaattgtg tttttaaaga gtgaaaacag 60

ttaggaaaca agtagaactg taatcagaac gctgcttcaa ttgatattaa aaataacctc 120

aataataatg taaaggttcc tttctcttgt gtcagttata ttcttaggga tagcctagaa 180

ggaatatatg gttagaacta agtgtgacta atcatctgag ccttgaagag aaacttcagt 240

gcctctaaac agatcatcta caaaacaaca ggtaaacatt tatgccagtt aagtgggtca 300

tgtttttgtt tcttgggttt ttcctaaatt taagtgaggt tgggcttacc ttgtagataa 360

aattatgttt tctttttggt aaatacttga atgtggataa cgtcaaatca gaatattttg 420

tgaggaggtg atgatttgaa attaagctag atttctaggg aggtgttggt tccaatgaag 480

gatgggaaga aattaaaata gtcttcaaac ttcttcctta ttatatttgg ttgctttgga 540

aaagattggt cctatcctca atctaattta ttcactatta atattttaaa aacattcctg 600

agatacttaa aaagacccac ttagcgatta tagttgctca atgaaacaag aatttattta 660

tgcatagatt tttctctgta tcttaccaaa atccacttta cttagataac actaaattgt 720

tcttaaagac tactcatttc ccaataatcc tttatgattt caaaatttct agtggctcag 780

aagtgaattt tattttattt gtctttcact tgaataaatg agaacccaga aattaataat 840

gttgtttatt gcttactgtc aggactattt caaagactaa gaagagtttc ttctaacccc 900

tccctctcaa aggaatccta aattattagt tgttagataa gttttgtatg ctaagatatt 960

caggtttata gtttatgtat gtgtgtatat atataaatat atatgtatat ataaatatta 1020

tgttcagttt ggagtctggc acaactccat tatgtggatt agagagtaag atattatgga 1080

tgataaagta ctaaatgaaa cataatattt atttataaaa gtgtgtagat tgttaaatca 1140

caaaaagagt gctatgacca ttatgtatga ggaaacaggc ctttgacctc ctggaaagca 1200

ctgctcaaaa gtcattagtg cccatttttg aattccccaa acagaaagct tcttagaaaa 1260

cacgctgaga ttttatttac agggaattct ttgacacatt tcaattggtg tgtagtcaag 1320

tatagcaagt acttaataat gactgaattt catgttccta cagtcataca tattcattag 1380

aagttttatg ttgttggtct gatctgattc ttctttgttt gtgggtggaa cggcactgag 1440

agaagtatag ttttttaaac ttgaacatgt tcagtagtta cattgcctta gaaaacccag 1500

acacatagca gtggaaatga aagaaatggc atcagaagtg acttaattta gcaattgtga 1560

ttcctcttgt aaaacaaaac aaaaaaacaa tgccatattt tttggagaaa agttggcaat 1620

ataggggttt cgttgtctgt ttcacaagaa gactcatttg ttcttttggg ggaaccagtg 1680

ccttacagat tttgtatata ctgtaattat tcaggactag ggaacaaaca attgtattgt 1740

atttgttaca gattgtatat ggctttgttt taacattccc ctaaataaaa tggcttcatt 1800

ctccccttgg aaaaaaacat gactgttatg ttataaaaca aaaaaaaaaa aaaaaaaaaa 1860

aaaggtgggg gtaccggggc aaaacgtgtc ccggggggaa tggtttcccg gcccacaaat 1920

cccccacatt gcgagaaaac cgtgcgaaca aaaaaaaaaa aaaaacgaaa aaaaaaaaca 1980

ggg 1983

<210> 110
<211> 758
<212> DNA

<213> Homo sapien

<400> 110
aaaaaaaacc acaaacaaga gaggattgat tgataatatg gggcatgctt aatctaatca 60

tgctcgagcg gcgcagtagt gatggatcga gcggccgccg ggcaggtacc taacatatag 120

tagacagtgg agagtggttc tctttcgttg tctcaggggc agacagatgg ggtgctggag 180

tcctctatca aagagtcaga gctctatccc agatgtgtaa tgaacgtggt cacagacata 240

ttgtcccatt accatttacc ttccctataa ccactgtgcc tccagccttg tagaatagac 300

acataggagc gcagcaatac gtctaaaaat aggagtgaga gagggcaggg catgcccgtt 360

cttgtggtag aagaaaagaa tgtcaaagaa agcagctggg actaatgaac tttacattag 420

ccatattcca ttatttcagc ttaagtcaaa tgtcggtcct catgaggcaa ctggctttga 480

caggagctac gctaatgtgc cacttaccaa cctttaattt ctgggtaaaa gcagaaagag 540

aaaaactaat ggatttttca ttttccagaa gagacaagaa tcaactacac tagtagtctg 600

tcagaacaaa agaaaacctg catccaatta caagaattat tactgtctct ttaataaata 660

accacattat taaaaaaaaa aaaaacaaaa aagggttggg ggtaccgggg ccaaggggtc 720

ccggggggaa ttgtttcggt ccatatccat acaaaaaa 758

<210> 111
<211> 3575
<212> DNA
<213> Homo sapien

<400> 111
atgaaattac aactcaggat taagagtctc actcaaaacc gcacaactac atggaaactg 60

aacaacctgc tcctgaatga ctactgggta aataagaaaa ttaaggcaga aataaataag 120

ttctttgaaa ccattgagaa caaagacaca atgtaccaga acacagctaa agcagtgttc 180

agagggaaat tcatagcact aaatacccac atcagaaatt gggaaatacc taaaatcaac 240

gtgctaacat cacaattaaa agaactagag aagcgagagc aaacacattc aaaacaagaa 300

ataactaaga tcatagcaga actgaaggag atagagacac aaaaagccct tcaaaaaatc 360

agtgattcca ggagctggtt ttttgaaaag attaacaaaa cagatagact gctagccaga 420

ataataaaga agaaaagaga gaagaatcag atagacacaa taaaaaatga taaaggggat 480

atcaccacta accccacaga aatacaaact gccatcagag aatgctatca acacctctac 540

ataaataaac tagaaaatct agaagaaata ggccgggcgc agtggctcac acctgtaatc 600

ccagcatttt gggaggccaa ggtgggcgga tcacctgagg tcaggagttc gagaccagcc 660

tagccaacat ggtgaaaccc cgtctctact aaaattataa aaaattagcc gggtgtagtg 720

```
gtacacgcct gtagtcccag ttacttggga ggctgaggca tgagaattgc ttgaacccag    780
gaagtggagg tggaggtgag ccgaaattgt gccactgtac tccagcctgc aacagagtga    840
gacactgtca cacaaaaaag aaagaaatat cacaatatgt cacaataggc cgggcgcagt    900
ggctcacacc tgcagtccca gcactttggg aggccaaggc agatggatca cctgaggtca    960
ggagtttgag accagcctgg ccaacgtgac aaaacccagt ctactaaaaa tacaaaaatt   1020
agccaggcgt gatggtgggc acctgtaatc ccagctactc aggaggctga gacatgagaa   1080
tcgcttgaac ccaggaggtg gagattgcac tgagctgaga tcctgccact gggctccagc   1140
ctgggtgaca gaatgagact ctgtctttaa aaaaaaaaaa aaaaaaaaaa aatcacaata   1200
agtcctaggg taaagatggg ggtacagaaa acaattaaat agaacaaaaa caactgtttc   1260
cttttcctgt gattcaagaa gggcttagat cttctactca gcatcctttt actaatgccc   1320
tccattggct ctcacgccca acatttcctt ttttatagct tattttgtaa tgcctcctta   1380
attatccttt aatagaagcc accgctgata agctacctac actcatacag aagcattaat   1440
ataatgcccc agatgtactg tttcagggca aaaaggaaaa taatttccaa caaagtggtg   1500
tgtgtctcac tgtcagatgc ttgcacttac acacggaatc gctgtgcatc cgacagaggc   1560
tgattggcac atggggcacg gggattgtca gctcaaacac cgtcagcagc gttgcccttg   1620
gaaatgggat ttcccagaac agtaaacgtg tctgtccttg atttacagag tagctacatt   1680
cctaggaaat ccagggtaca ttaaaactca ccatgttacc caggctggtc tcgaactcca   1740
ggcctcaagc aatcctccca catcagcttc ccagaatttt gggattacag gcatgagcca   1800
ccacacccag ccagaatatt ttatttctgt tagacacaga gcgttcgttg actcgtctgg   1860
gcgttagtgt taatattctg tacttgaagc aagcccacca agcggctgaa ctgggtggat   1920
aatggaaaat gtcctgtgga tttgggagtg agacaaaccg gcttgagtct aacctctcag   1980
ttagtctaag gctccaagct tgaaagggtt aaatgaagta ctatatttgt tttgtttcgt   2040
tttcgttttg tttgaggctt tgctctgttg cccaggctgg agtgtagtgg cacaatctct   2100
gctcactgca acctccatct cccaggttca ggcgattctc ttgcctcagc ctccagagta   2160
gctgggatta caggtgcccg ccaccacacc cggctaagtt ttttttggta tttttagtag   2220
acacagggtt tcaccatgtt ggctagactg gtctcgaact cctgacctca agtgatccac   2280
ctgccttggc ctcccaaagt gctgggagta tgggtggtga gccaccacgc ctggcctaaa   2340
tgaagtacca catgaccgac cgaccgacct ggggaacata gcaagacccc atctctacaa   2400
aaatgtaaaa aataaaaatt agccgggtgt agtggtacat gcctgtaatc ctagatactc   2460
```

gggaggctaa ggcagaagga tcacttgagc ccaggagttc gaggctgcag tgagctgtga 2520

tcgtgccact gcactccatc ctgggtggca gagtgaggcc ctgtctcaaa ataaataatc 2580

cagtcccccc caagaaagga atgaagtgct ataatgagaa aaatcctagt acctaacata 2640

tagtagacag tggagagtgg ttctctttcg tttctcaggg gcagacagat ggggtgctgg 2700

agtcctctat caaagagtca gagctctatc ccagatgtgt aatgaacgtg gtcacagaca 2760

tattgtccca ttaccattta ccttccctat aaccactgtg cctccagcct tgtagaatag 2820

acacatagga gcgcagcaat acgtctaaaa ataggagtga gagagggcag ggcatgcccg 2880

ttcttgtggt agaagaaaag aatgtcaaag aaagcagctg ggactaatga actttacatt 2940

agccatattc cattatttca gcttaagtca aatgtcggtc ctcatgaggc aactggcttt 3000

gacaggagct acgctaatta ccacttacca acctttaatt tctgggtaaa agcaaaagag 3060

aaaaactaat ggatttttca ttttccagag agacaagaat aaaataatag tagtctgtag 3120

aaaaaagaaa acctgcatca attacaagaa ttattaatgt atctttaata aataaccaca 3180

ttatttagct gtttaatttc ctaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 3240

aaaaaaaaca aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aacaaaaaaa 3300

aggagggggg gggggcgaga aaaagagccg agggggggagc acagagcggg ccgccgcgca 3360

catatgaaaa aagcgaccca gaagaagaaa cacaaaacca gcaagcgcaa acagaagaaa 3420

taagaaagag aaaaagttac gagacgaata gaaaggaaat aactacagga ccaacacggg 3480

acaaaccaaa agcaaataaa caaagaaaat aagacagaca caagatgcca acgagctaac 3540

gcccggacaa tggaaacagg taaacaacat aaagc 3575

<210> 112
<211> 442
<212> DNA
<213> Homo sapien

<400> 112
actgagcagc tacggaagtg caaggcactg taggagtagg gtgagtatac tccccacaag 60

ggctcagggt caggcagggg acggtagaga taaaaaccca cagaccatac acatagctgg 120

cactgtctct gagggttttg tgaggcacac aaatgcttag gagactagac gaagtaagac 180

aatgtctttg acatgaggca gaaatcaacg gaaagcatgc gcttttagaa catgtgtggg 240

actgtttttt ggtatcagca gactgaagag gctttttaaa cgtggaggga aggcaaactg 300

aggcatagag atgccaatac caggtcttgt caggaagaac agagtccaat ttggctgcag 360

gatagggcat atgtagggga ggggataaga ctggcatggg ggcagagggg gacttgaatg 420

tcaggtgaca gagtcaaagc tt 442

<210> 113
<211> 412
<212> DNA
<213> Homo sapien

<400> 113
tgtcatacta taaggcgaac tgggcctcta gatgcaattg ctcgagcggc gcaggtgatg 60

gatgttcgcg gcgaggtatc agaagctgtg atgtctgcct tgtagtcctg tgcttgttac 120

tgtaattttt tttttttttt tacgaagcac gtgactggac taatgtaagg cagatgacgt 180

gatctttaag actgctatat atatcagtct cttactctat aaggttttaa attagaaaag 240

gcttatatgg ttaactacct tagactatat ctacagcagg gtctggtttg ccagaacaag 300

tttaaagtgg ctgtttatta agttggctat tttcagaatt gaaactataa gaccgccatt 360

tgacactgaa acttgcgtga atcctaaatt gcatcaatta tctatttgat aa 412

<210> 114
<211> 625
<212> DNA
<213> Homo sapien

<400> 114
gcaacaacaa ctgaatggct gtaatacatt aatgtataca gatgaattga gaagtcttct 60

agtgaaatgg ctcagatctt tgttcttggt ccagtcctgt tcagtttttg atcagtgcct 120

tgcaatatca cttgatcgac tcacttaaca tttatacaag agtgcagagg cctcctcaga 180

gaatggatgg tagaaatgca ttgatgagag aacgtttatc tatctatctg tctatctatc 240

tatctatcta tctgtctcta tctaagaagt cataaaggct gagtctaata aggcaaaaaa 300

aaaaagaaaa aaaaaaaaag ctgtggcgat acccagggcc aaagcgtgat cccggcgcca 360

actgcgaaat ccgctcacaa tcccaacaac acccccacaa cccccccccc agccccaaca 420

ccaacacctc aacaaaacct cacaacaccc ccaccacccc acacagccac ccccctacca 480

cacaaccaca tcacaccacc accgccccac ccacaccaca caccccacac acactccgaa 540

caccaccgcc cactccacac acccaccaac caccagcacc aacacaccca cacacccaca 600

ccgccacccc cacaccacac gcagc 625

<210> 115
<211> 378
<212> DNA
<213> Homo sapien

<400> 115

gcggcgcccg ggcaggtaca tagtgcagat gcagtatata atttcaggct aggaaaatta 60

gctactagta tgtatctgac agttcctaat agctaagagg cctaagaatg cagacgggga 120

gaaaaaaaac caaaaccaaa aaaaaagaca cctctccaat tgctgggagg gcctgggaat 180

aggtgaagat caaaccacag tgggagagga gggtaaagat gtgagcttca agcgggtaat 240

gggcaagcca cacctcccag ttcctaggag ggaatcgcca cggccgactt cagcattctc 300

gtctttacta agacttaccc atagagaact acagcaggaa accgatttct tcattcattc 360

tctttaaaaa gtatgaat 378

<210> 116
<211> 8905
<212> DNA
<213> Homo sapien

<400> 116
atggcggcgg cgctggggcc cccagaagtg atcgctcagc tggagaacgc ggctaaagtt 60

ctgatggtga ggacgccgcg cccctcagac cccgggattc gcgggccccc ggtcggccct 120

gccactccag gccttgctgc tcgctgggct ggcgactggc aagggcctgc agggagcctg 180

gaagtggagg aggaggtggc ggtggcgtgg cgcaggattc ttcagcctac tttcctcctg 240

ccgtcgtccc ctccttccag gagctgtccc cttcccctgg ctgcccagca ccccagtcgg 300

gcgtgggaat atagtggtgt agcaaagaga atttcttcac cttacaccct gccccacaga 360

ctgggtcgca gagcaaggcg ccgggaagga gttggggtta tccccgcagg gcttcgggcc 420

tctcatatac tagtccttct gtctggaatg cttttcttcc ctgtcacttc atccttcagt 480

tctctcagta gtcagtttct cagggaagcc ttccttagcc tgcctgaaag tataccctgg 540

gtgatagatt ggattggatt ggattggatc ggatcggatc ggatcggatt ggattatatt 600

gtatttattt ttaagagaca gggcagctgt caaaatggaa gttcagggtc actagaggtt 660

ggcacatgtc tccagggtaa acacatgagt gcttgcattc atctttggat ccctgcgttc 720

gcttctgttt tagcttttga tgattcctta atttcttctg ccacagccat aatggaagca 780

gttgtccgag agtggattct cttggaaaaa ggtagcatcg agtctctgcg aacattcctt 840

ttaacctatg tcttacaaag gcccaacctt caaaagtatg ttcgggaaca gattctacta 900

gcagtagcag taattgtaaa aagaggatca ttagataaat caattgactg caaaagcatt 960

tttcatgaag tcagccagtt gattagtagt ggcaatccca ctgtgcaaac tctggcctgt 1020

tctattctga ctgcgctatt gagtgaattt tcaagttcaa gtaaaactag caacattgga 1080

ttgagcatgg aattccatgg taactgcaaa aagagttttt caggaagaag accttcgtca 1140

gatcttcatg ttaactgttg aagttctgca ggagttcagc aggcgggaaa acctcaatgc 1200

tcagatgtct tcagtatttc agcgttacct tgcactcgcc aatcaagtct tgagctggaa 1260

ctttcttcct ccaaatttgg gcagacatta tatagctatg tttgaatcct cgcaaaatgt 1320

gctgttgaag ccaacagagt cctgcgggag actcttctgg acagcagagt tatggagctt 1380

ttcttcacag tacatcgaaa aatccgagaa gcattcagat atggcaccaa gattctctgc 1440

agtgccttgc ccagttagct tctcttcatg gacccatctt cccagatgaa ggatcacaag 1500

ttgattatct agcacacttc attgagggat tactgaatac tatcaatgga attgaaatag 1560

aagattctga agctgtgggg atctccagca ttatcagcaa cctgataacc gtgttcccac 1620

gaaatgtttt aactgccatt ccaagtgaac ttttctcctc ctttgttaac tgcctcacac 1680

acctcacttg ttcttttggg cgaagtgctg cattggaaga agtgcttgat aaagatgaca 1740

tggtatacat ggaagcatat gataaattgt tggagtcctg gttaactttg gttcaagatg 1800

acaaacattt ccataaaggc ttttttaccc aacatgcagt tcaagttttc aattcctata 1860

ttcagtgcca cctagctgct ccagatggca caagaaattt gactgccaat ggtgtggcct 1920

ctcgtgagga ggaagaaata agtgaacttc aagaggatga tcgagaccag ttttctgatc 1980

aactggccag tgtaggaatg ctaggaagaa ttgctgcaga acactgtata cctcttctga 2040

caagtttatt agaagaaaga gtaacaagac tccatggtca gttacaacga catcagcaac 2100

agttacttgc ttcaccgggt tcaagcactg ttgacaacaa aatgcttgat gatctctatg 2160

aagatattca ctggcttatt ttagttacag gctacctctt agctgatgat actcagggag 2220

agactccgct aatacctcca gaaataatgg aatattccat taagcattca tctgaagttg 2280

acattaatac aacacttcaa attttgggat ctccaggaga aaaggcttct tccatcccag 2340

ggtacaacag aacagattct gtgattaggc tgttgtctgc cattctcaga gtttcagaag 2400

ttgaatctcg agcaataaga gcagatctca ctcatctact aagtccccag atgggcaaag 2460

atattgtttg gtttttaaaa cgctgggcaa agacttatct cctggtggat gaaaaactgt 2520

atgatcagat aagtctgcca ttcagtacag cgttcggagc agatacagag ggttctcagt 2580

ggataattgg ctacctctta caaaaagtca tcagtaacct ctcagtctgg agtagtgagc 2640

aggaccttgc aaatgacact gtgcagctcc ttgtcacttt ggtggaaaga agagaaaggg 2700

caaacttagt aattcaatgt gagaactggt ggaatttagc taagcagttt gcaagccgaa 2760

gcccacctct taatttcttg tcaagtcctg tgcagaggac attgatgaag gctctagtct 2820

taggaggttt tgcacatatg gacacagaaa ccaaacagca gtattggaca gaggttcttc 2880

agccacttca gcagcgattc ttaagagtga taaaccaaga aaacttccag cagatgtgtc 2940

agcaagagga agtcaagcag gaaatcactg ccacactaga ggccctgtgt ggcattgctg 3000

aggctaccca gattgacaac gtagcaatcc tgtttaattt tttaatggac ttccttacca 3060

attgcattgg attgatggaa gtttacaaga ataccccaga gactgtcaat ctcattatag 3120

aagtttttgt tgaagttgca cataaacaga tatgctatct tggagagtcc aaagctatga 3180

acttatatga agcctgcctt actttgttgc aagtgtattc taagaataat ttagggcggc 3240

aaagaataga tgttacagca gaagaagagc aataccaaga cctgcttctc attatggaac 3300

ttcttactaa cctgctgtca aaagaattca tagatttcag tgatacagat gaagtgttta 3360

gaggacatga gccaggtcaa gcagcaaaca gatctgtgtc agcagcggat gttgtgttgt 3420

atggagtaaa cctaattctg cccttgatgt cacaggatct cttgaagttt ccaacccttt 3480

gtaatcagta ctacaaatta atcacattta tctgtgagat tttttcctgaa aaaataccac 3540

agcttcctga ggatctgttt aaaagtctga tgtactccct agaattagga atgacatcaa 3600

tgagttcgga ggtttgccag ctttgcctgg aggccttgac accgttagct gaacagtgtg 3660

caaaagcaca agaaacagac tcaccacttt ttctagcaac acggcacttt cttaagctgg 3720

tttttgatat gctggttttg caaaagcaca acacagagat gaccactgcg gctggcgaag 3780

ctttctacac gttggtgtgt ttgcaccagg ctgaatattc tgaactggtc gaaacattac 3840

tatcaagtca gcaagaccca gttatttacc agagattagc agatgccttc aacaagctca 3900

ctgcaagcag cactcctcct acgctggatc ggaagcagaa gatggccttc ttaaagagtt 3960

tagaagaatt tatggcaaat gttggtggtc tcctttgtgt aaaataaaca acagaacttt 4020

atgcttaatt tagatccttt ctgcaaagtg cactgaattg ctgaaagttg acttgagtct 4080

tgtcctattc ctcagttcat ttggccattt tggattttgg agagcctgaa actttgatat 4140

gtatgtaata cagtgaaaca ggagaggtca acttggcatc agcttctgct gttaagtgtt 4200

agccacaatc tgtcatatat atgtctttta gattctgaat ggtgatttaa aattttcaaa 4260

atgaaattcc atatatgtgc aaacagatat gggcaccacg aaatacatat gcagtgcctt 4320

ttttcctttt aacataggtg gctagccaaa gtttagaatt tttgtcatta aatatgaaat 4380

ggatatatgc taggcagtgt ttctcaaaat ctccacagat cgcctgcatc acttgaggag 4440

ctggtgaaaa ggcagattct taggcccaac tgtagacctt cagagtcaga atgtctggtt 4500

gttgggccca ggagtcttca tgttaataag cttctccctt tcgtcacccc aaaagttttg 4560

aatcaatgaa agagacattg aaaactctta agaggttttg tgctttctag cttttcctcc 4620

ctttgatgat tgggttttat aattcagcag gaaggggaaa catcatcagg ggtttgttgg 4680

cttttttctta gcttgctttc ttgcttgctt gctttcttgc ttttcttgct ttctgtctct 4740
ctctttcttt tctctctctc tctcacatca acccagtgct gcaggttttg tgtaatacaa 4800
gtcactaatc atactctgat gcctgaactt gaggaggaaa atacatgtat atttttgttc 4860
cgtaaaaata accttaggaa ctgtagccat ttcattgcct taattttaag aggaaaatac 4920
aaaaacagct gatttgtttt agtaagaaac cacgtcttga tgcttcagag ttggtttagg 4980
gtgttagctg ctatgaacct gttgcccctt tcgatcgtgt atttatgtag gtttatcagt 5040
gaaatgaaag gcttgtttcc gtctagtcta actttttgag tgtgtttcta tccagccaca 5100
tagcccatat ctactctaaa tggcttgctt aagcaataat tattttaaag gatgtgaatc 5160
actgattcac acagactatt gcacgttggg gcattagggg caataattct tatccagaca 5220
tgggagccag tgaatttaat ttcagagatt aaaaattcac tttagatcct ctagtttgat 5280
ctcttaatca ggatttttat acagctgcca ggctccccta attcagtgtg ccagcttaca 5340
atgtggaaat gaaagctaat ttatacacag caggcatatg aaactccact cattgcagta 5400
ctttcacagc acagtgacag gtagaggact ctggcacagg tgcactcatg aaactctgct 5460
tccaccatgt tcctgacacc tatctattaa accattctgc aaatacggtt tttctacctg 5520
attgcatata gcatatgtgt cattacatgt gatgctgtgc aaaactttgt ataattctgt 5580
gttattaaca gttaacaaaa ctggagcatc tgaattacat ccaacctgtg catgtgatgt 5640
taggtagatg tgaatgcagg gccttgggcc ataacttaca tttctctcaa tttgattagc 5700
tttgagtcac aattaagggg aagcaaaaac atcttgaaaa gactgctagg aaggaaatta 5760
atatcagtca tccagaagta cacgtttctg tattttaaaa aatactttga tgcatttatt 5820
tttaggtgtt tttttttcc ccttaaaaaa cttgaagtga tatgcagcag taatctattt 5880
gttttgcatt gttcttggtg ttttgtgttt cccagatccc tcaagctttc tcagctgttg 5940
cgaattatgt gtatctgtgt gtgtgctaag tacagtctct ttaccaaagg gcactgaaac 6000
acacaattga ctggacaggt ccacgcgcca tgacaaaact ataatcaagt tattaaaact 6060
aaagaggagt gggaaaggaa tgccttggta agtaaaaagg catctatatt taataacttt 6120
tatccagatg gcaacatatt tgcaaaattt gcccagatcc tattacaata ctaaaaatag 6180
aaaatttcac ctccatattc ctgaggtgta atttcattag actagtttta gtttaaaaag 6240
accttcttca gattggacca aataatactt ataagatcag cagaatgttg aatattagct 6300
cactggggtg gggagaagcc actaccattt tttaggtgat ggggatgcca ctgagttgca 6360
acggctagac cttttcaggg tggttgtgtc catgtttgcc tgattggatg cttattcact 6420
ttgtgttttc ttttgtttta ttttgtccaa ttttgtcttt agctgtgttt attaacttct 6480

ccggtcttgt tttgttttaa tgctcttggc ccagtgggtg tcaagaacac tggcttaatt 6540
caagtcagtt gatttttttt ctattaaaac tgttgttaaa atatttttta aaacaaaaac 6600
attatttgtg ccctctttta tatatgtcaa agggacactg tcaagtattt catttttaga 6660
tttttgtttt ataaaatttc tgttgttcat atagtatcct ttaacctcta gttttccata 6720
catcctttgt ttgtttctca ttttattttc cttgacccat ttatttccca aggcacaatc 6780
actaaagact ttgtactttc acagtctgtt aatgtggtag cacctgtaac tgtgttcttg 6840
ttctgttaaa aggattgatt tgcttttata gtccttgtgc tggatgagtg gctgcctcag 6900
tagcaaaact acctgacagt atttgacagt gtcctttcca gcaccattat ttgggtcttt 6960
cagggtggcc atctctgtta gaagacagta gcatgttaac atcactgcat tgagtttttg 7020
tctggtgtaa agtatgactt ttaatgtaaa caaactgcag gtttttttca aactaatttt 7080
aagaatttag tcttatttcg ttgtaaactg tgtatctaat tatattacat tactctgttc 7140
agatgggatg gttactacca cttgtccatg attttcattt gaaaagcaag tatctatatc 7200
atttcccccc agtcagcatt atttaacact ccccttaact gtctttgaac tttctctttt 7260
aacaaaaatg tcaagtcttt acagttgtaa tatcaccatg tttcccattt ctgttaatac 7320
ttctatgaac ccctaaagta ttgaagggaa ctagctgtca gtttcaagga ttacaagttt 7380
gagtctccta gtattcaaca tcattctgaa ccctgaaata atatttttct ctgttaaaca 7440
atttttatct gtttgccacc tctgttgtta gaggtggttg tcaattgacc ttactaagtt 7500
agctgtcttt gatgaggaat tattgttatt ggttcctgaa taaaacatta acctttttaag 7560
tcagaaggaa cctcggtact tcttaaggtt tgtttgtgtt ttctaaaacc agagaataag 7620
gaactgattt ggctatgagg tttaacatta taattttctg taagctttcc cacaaaaaaa 7680
cattgttgat ttgaggatat aataatgttt taatcttttt aaaatataag tggttattct 7740
ctgacttggt aactatgttc tgaaaacact gcatttaaga atttttaaaa attggttttc 7800
taaaattaaa atgtccaaat taggcatatt gctgagctca aattgatgtg aaatgccatg 7860
gttccagttg aattttaagc atattttcat ttagatataa aatatatgaa gtatgctttg 7920
ttgattatag tgagaaccca tgacatagtt aaccaaagaa tatgtttggt tcaaataaaa 7980
atagaagctt aatactgggc attcatactt tttaaagaga atgaatgaag aaatcggttt 8040
cctgctgtag ttctctatgg gtaagtctta gtaaagacga gaatgctgaa gtcggccgtg 8100
gcgattccct cctaggaact gggaggtgtg gcttgcccat tacccgcttg aagctcacat 8160
ctttaccctc ctctcccact gtggtttgat cttcacctat tcccaggccc tcccagcaat 8220

tggagaggtg tcttttttttt ttggttttgg tttttttttct ccccgtctgc attcttaggc 8280

ctcttagcta ttaggaactg tcagatacat actagtagct aattttccta gcctgaaatt 8340

atatactgca tctgcactat gtacctacta gggatctgac ctcaagtgtt ttctgagccc 8400

aggcttcctg gtgtggtgtc ttttaccaca taaaattatt acaaattgca aatgttggta 8460

ttgtgatttg attatctgta caaagaaaga agctctatgc agtgagtttg tggtttaatg 8520

gtcacaaaaa tgttagcact gctaccactc agcacgtgta aaatttttta aatttataaa 8580

tattaaaatt ttaaacttac actaagactt ttcagtttta tttaaagacc cagggatgag 8640

tgtactgttt aaatatttac ctctattaac ataactaatg aaggtataaa attgcattta 8700

gtttttcaga agatgctgca atatgatttt aggaaataag gctatgtatt gagccagtta 8760

taggctgaat atcaggttga taaaatttta tttgtatttt taaaattcat aaatgggagt 8820

taaaatgtgt cttttcacta aatattttta ttacaaaaaa aaaaaaaaaa aaaaaaaaaa 8880

aaaaaaaaaa aaaaaaactg cggcc 8905

<210> 117
<211> 827
<212> DNA
<213> Homo sapien

<400> 117
tcgcggccga ggtaccctgc atcactgcca tggttgtgct attctcatct caacatagaa 60

ttggtgggtt ctcctaaggg tgtcaggaac ctctaaaaag atgtgattct ttgggagggg 120

atatttgaaa ttccaacttc cattccccct agcaaaagga agcagctgct gtttaagggt 180

tttatctgag ccactttaaa gatgaatcca tggtattact ctggatacta gccattcctt 240

aggattttaa ggtcacattt tattcctgga tgctttatgt ccccacctcc acctgagccc 300

tcatcctctg ttccctacta tactcccaac ttctactctt tgttttatcc acctatccct 360

attacctgac cctttgtctt ccctgtctcc catccttggg gggacatgca gccctgtggt 420

catggttctg atgacatcat cagggcagcc ctcctgccca ggtattatgg cctgtcagca 480

ttccctgtgc cctccaaacc ttaggcctag aatgcggagc tgccaacata acattcaccc 540

ttttgaacag atggagtcag gcacactaac acagccttct gtcctcaata acacagccat 600

tattgccact tggctcagtc gtcaatgtaa accctcagag tcagctgaac tattttaggc 660

caaacatact gtttttgtaa agtatttttc attaataaat ctataagaca gttctattta 720

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aggctggggc gaaccggggc caacggctcc 780

cgggggaaat tgtttcccgc caaattcccc caaaaaatgg caaaacg 827

<210> 118
<211> 6470
<212> DNA
<213> Homo sapien

<400> 118
ggctccctgg tagctatagc agccgcggcg gttaagtatg cggcgccagg agctgctaaa 60

tgtgaacaat aatgtcttgg aagagaaatt atttttcagg gggtcgtggt agtgtacaag 120

ggatgtttgc acctcgaagc tcaacctcca tagcccccag caaaggcctc agcaatgagc 180

cagggcaaaa cagctgcttc ctcaacagtg ccctgcaggt tttgtggcac ttggatatct 240

tccgacgtag ctttaggcag cttacaactc acaagtgcat gggagattcc tgcatctttt 300

gcgctctcaa gggaatcttt aaccagtttc agtgtagtag tgaaaaagtg cttccatctg 360

acactctccg cagtgctctg gcaaagactt tccaggatga acaacgtttc cagctgggaa 420

ttatggatga tgctgcagag tgctttgaaa acctcctgat gagaattcac ttccacattg 480

ctgatgaaac caaagaggat atatgtactg cccaacactg catttcccat cagaaatttg 540

caatgacatt gtttgagcag tgtgtatgta ctagctgtgg tgccacttct gatccgctgc 600

ctttcatcca gatggtacat tatatctcca ccacttccct ttgcaatcag gctatttgta 660

tgctggaaag acgagagaaa ccttcaccaa gcatgtttgg tgagctgctg cagaatgcca 720

gcaccatggg ggatctgcgg aactgtccaa gcaactgtgg agagaggatc aggattcgcc 780

gtgtgttgat gaatgctcca cagattatca cgattgggct ggtatgggac tcagaccact 840

cagacttagc agaagatgtt atccacagcc tgggaacctg ccttaagctg ggtgatctgt 900

ttttcagagt gacggatgac cgggccaagc aatctgaact gtacttagtt ggaatgatct 960

gttactatgg caaacattat tctacattct tttttcaaac aaagattcgc aaatggatgt 1020

attttgatga tgctcatgtc aaggagattg ggcccaaatg gaaggatgtg gtgaccaaat 1080

gcatcaaggg gcattatcag cccctgctgc tgctttatgc agatccccag ggtaccccag 1140

tttccaccca ggacctgcct ccccaagctg agttccagtc atacagcagg acatgctacg 1200

acagtgaaga ttcaggacac ctgactgata gtgaatgtaa tcagaaacac acatccaaga 1260

aagggtcact gatagagcgc aagaggagct ctggtcgggt taggaggaaa ggcgatgagc 1320

cccaggcctc gggataccac agtgaaggag aaacactgaa agagaagcag gctcctagaa 1380

atgcctccaa accatccagc agcaccaaca ggctgagaga ttttaaagag acagtcagca 1440

atatgatcca taacagacca tccctggctt ctcagaccaa tgtaggctct cactgcaggg 1500

gcagaggagg agaccagcct gacaaaaaac ctcctaggac cctgccttta cactctcgtg 1560

actgggaaat agagagtacc agcagtgagt caaaatccag ttcttccagc aagtatcgtc 1620
ccacatggag acccaaacga gaatctctga atattgacag tatctttagt aaggacaaaa 1680
ggaagcactg tggctatacc cagcttagcc ccttttctga ggattcagct aaagaattta 1740
taccagatga accaagcaag ccaccttctt acgacattaa atttggtgga ccaagccccc 1800
agtacaagcg ctggggccca gcacggccag gctctcacct tttagagcag cacccccgac 1860
taatccagcg aatggaatct ggctatgaaa gcagtgagag gaacagcagc agccctgtca 1920
gcctggatgc agccctgcct gagagctcaa atgtctacag ggatccaagt gctaagagat 1980
cagctgggtt ggttccttcc tggcgtcata tcccaaagtc gcacagcagt agcatcctgg 2040
aggtagactc cacagcatcc atgggtggct ggacaaagag tcagcctttc tctggtgagg 2100
agatatcttc taaaagtgaa ctggatgaat tgcaggaaga ggtggccagg agggcgcagg 2160
aacaggaact tcgaagaaaa cgggagaagg agttagaggc agcgaaaggg tttaaccctc 2220
atcctagccg cttcatggac ttggatgaac tgcagaatca ggggaggagt gacggctttg 2280
agaggtccct gcaagaggca gagtcagtgt ttgaagagtc actacatctg gaacagaaag 2340
gagactgtgc tgcagctttg gctctctgta atgaagctat ctctaaacta agacttgccc 2400
tgcatggtgc cagctgtagc acgcacagca gagccctagt cgataagaag ttgcaaatca 2460
gtattcgaaa agcacggagc ctgcaggatc gcatgcagca gcagcaatca ccacagcagc 2520
cgtcgcagcc ctcagcctgc ctcccaacac aggcggggac tctctctcag ccaacaagtg 2580
aacagcctat cccgctccaa gtattgttaa gccaagaggc ccaactggaa tccggcatgg 2640
atacagagtt tggggccagt tctttcttcc attcacctgc ttcctgccat gagtcacact 2700
catcactatc tccagagtca tctgccccac agcacagctc ccccagtaga tctgccttga 2760
agcttctgac ttcggttgaa gtagacaaca ttgaaccctc tgcattccac aggcaaggtt 2820
tacctaaagc accagggtgg actgagaaga attctcatca tagttgggag ccattggatg 2880
ccccagaggg taagctgcaa ggctctaggt gtgacaacag cagttgcagc aagctccctc 2940
cacaagaagg aagaggcatt gctcaagaac agctgttcca agaaaagaag gatcctgcta 3000
acccctcccc ggtgatgcct ggaatagcca cctctgagag gggtgatgaa cacagcctag 3060
gctgtagtcc ttcaaattca tcagctcagc ccagccttcc cctgtataga acctgccacc 3120
ccataatgcc tgttgcttct tcatttgtgc ttcactgtcc tgatcctgtg cagaaaacta 3180
accaatgcct ccaaggccaa agcctcaaaa cttcattgac tttaaaagtg gacagaggca 3240
gtgaggagac ctataggcca gagtttccca gcacaaaggg gcttgtccgt tctctggctg 3300
agcagttcca gaggatgcag ggtgtctcca tgagggatag tacaggtttc aaggatagaa 3360

gtttgtcagg tagtctaagg aagaactctt ccccttctga ttctaagcct cctttctcac 3420

agggtcaaga gaaaggccac tggccatggg caaagcaaca atcctctctg gagggtgggg 3480

atagaccact ttcctgggaa gagtccactg aacattcttc tcttgcctta aactctgggc 3540

tgcctaatgg tgaaacttct agcggaggac agcccaggtt ggcagagcca gacatatacc 3600

aagagaagct gtcccaagtg agagatgtta ggtctaagga tctgggcagc agtactgact 3660

tggggacttc cttgcctttg gattcctggg tgaatatcac aaggttctgt gattctcagc 3720

ttaagcatgg ggcacctagg ccaggaatga agtcctcccc tcatgattcc catacgtgtg 3780

taacctatcc agagagaaat cacatccttt tgcatccaca ttggaaccaa gacacagagc 3840

aggagacctc agaattggag tctctgtatc aggccagtct tcaggcttct caagctggct 3900

gttctggatg gggcagcag gataccgcct ggcacccact tagccaaaca ggctctgcag 3960

atggcatggg gaggaggttg cactcagccc atgatcctgg tctctcaaag acttcaacag 4020

cagaaatgga gcatggtctc catgaagcca gaacagtgcg tacttctcag gctacacctt 4080

gccgaggcct cagcagggag tgtggggagg atgagcagta cagtgcagag aatttacgtc 4140

gcatctcacg cagtctcagt ggcaccgttg tctcagagag ggaggaagct ccggtttctt 4200

cccacagttt tgattcatca aacgtgagga agcctttgga aaccgggcac cgttgttcca 4260

gctcctcttc cctccctgtc atccatgacc cttctgtgtt tctcctcggt ccccaactct 4320

accttcccca accacagttc ctgtccccag atgtcctgat gcccaccatg gcaggggagc 4380

ccaatagact cccaggaact tcaaggagtg tccagcagtt tctggctatg tgtgacaggg 4440

gtgaaacttc ccaaggggcc aagtacacag gaaggacttt gaactaccag agcctccccc 4500

atcgctccag aacagacaac tcctgggcac cctggtcaga gaccaaccag catattggga 4560

ccagattcct gactactcca gggtgcaatc ctcaactaac ctacactgcc acactaccag 4620

aaagaagcaa gggccttcag gttcctcaca ctcagtcctg gagtgatctt ttccattcac 4680

cctcccaccc tcccattgtt catcctgtgt acccaccatc tagcagtctt catgtacccc 4740

tgaggtcagc ttggaattca gatcctgttc cagggtcccg aacccctggt cctcgaagag 4800

tagatatgcc cccagatgat gactggaggc aaagcagtta tgcctcccac tctggacaca 4860

ggagaacagt gggagagggg tttctgtttg ttctatcaga tgctcccaga agagagcaga 4920

tcagggctag agtcctgcag cacagtcaat ggtaaaggtt attcctttcc tttcctggag 4980

ctacaccttt ctttgtaaaa ctgtactgtg ggccgggcgc ggtggctcac acctgtaatc 5040

ccagcacttt gggaggctga ggcgggtgga tcacgaggtc aggagattga gaccatcctg 5100

gccaacatgg tgaaaccccg tctctaccaa aatacaaaaa attagccagg cgtgacggtg 5160

cgtgcctgta gtcccaacta ctcggaaggc tgaggcagga gaattgcttg aacccgggag 5220

gcagaggttg cagtgagccg agatcgcacc actgcactcc agcttggcaa tagagtgaga 5280

ctccatctca aaaaacaaaa caaaacaaca acaaaataaa ctactgtggc agcgttggta 5340

ccctgcatca ctgccatggt tgtgctattc tcatctcaac atagaattgg tgggttctcc 5400

taagggtgtc aggaacctct aaaaagatgt gattctttgg gaggggatat ttgaaattcc 5460

aacttccatt ccccctagca aaaggaagca gctgctgttt aagggtttta tctgagccac 5520

tttaaagatg aatccatggt attactctgg atactagcca ttccttagga ttttaaggtc 5580

acattttatt cctggatgct ttatgtcccc acctccacct gagccctcat cctctgttcc 5640

ctactatact cccaacttct actctttgtt ttatccacct atccctatta cctgaccctt 5700

tgtcttccct gtctcccatc cttgggggga catgtagccc tgtggtcatg gttctgatga 5760

catcatcagg gcagcccccc tgcccaggta ttatggcctg tcagcattcc ctgtgccctc 5820

caaaccttag gcctagaatg cggagctgcc aacataacat tcacccttt gaacagatgg 5880

agtcaggcac actaacacag ccttctgtcc tcaataacac agccattatt gccacttgct 5940

cagtcgtcaa tgtaaaccct cagagtcagc tgaactattt taggccaaac atactgtttt 6000

tgtaaagtat ttttcattaa taaatctata agacagttct atttaaaaaa aaaaaaaaaa 6060

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaacaaaaaa aaaaaaaaaa aaaaaaaaaa 6120

aaacaggtgg gggccgcgcg cgggcgcgcc cccgagagaa aatttccaca aacacccgtg 6180

gggggggcggc gggcgcccca agtgagtgac agagaaagaa agacgaggag cacaacagga 6240

ggtccgtcct ccagaagaaa acaaccgcgt gcggcacaga acaaaggagg tggcgggggg 6300

tgcgctccac cacgacaaat aagaaaaccc cgcgggggggg ggaaaaacag cagacgagtg 6360

tcgtgaagaa caacaaccca caggagagag gcctcgtgga caaggcacac agggggtgct 6420

cacaaaacaa gggggtacaa agaaggagac gcaagaaaac ataattgccc 6470

<210> 119
<211> 435
<212> DNA
<213> Homo sapien

<400> 119
gtataatcat ataggcgcat ggttctctaa tgctgctcga gcggcgcgtg tgatggatgc 60

gtggcgcggc gaggtacctc tcaacactga gaactgtagt agttgaaacc actgttctag 120

tgggcagtta gaacagttgt tttccccgtc ttgttcccca cagagctgcc caagttatta 180

tctgctcctg gggttggacc atctgtttta tgacagttat gatattgttg tttaaaaaaa 240

atccaaattg ttactttgat ttatatgatc taactctgaa tcacggaagt attactatga 300

tgttcaaaac tctgattgac tctacttgct ttaaaaactc tcagatccct tctgcattta 360

tcatcagaga tcggtaaaga tgacaacaag caggtctaaa gttctgagat gttagcacat 420

acccttttca caatt 435

<210> 120
<211> 1262
<212> DNA
<213> Homo sapien

<400> 120
ggccgagttt tttttttttt tttttttttt tgtttttttt tttttttttt tttttttttt 60

tgttttagat atgttgcttt tattcaaaag aataaaatgc ttgacaaact ctttaatcac 120

aaggtttgaa ccaaaccacc agtcttctac aacaactctg tgaggtaggt atctgcatag 180

ccacaaggga tccacatagt cctttcttcc cttgtacctc tcaaacactg agaattgtag 240

tagttgaaac cactgttcta gtgggcagtt agaacagttg ttttccccgt cttgttcccc 300

acagagctgc ccaagttatt atctgctcct ggggttggac catctgtttt atgacagtta 360

tgatattgtt gtttaaaaaa aatccaaatt gttactttga tttatatgat ctaactctga 420

atcacggaag tattactatg atgttcaaaa ctctgattga ctctacttgc tttaaaaact 480

ctcagatccc ttctgcattt atcatcagag atcggtaaag atgacaacaa gcaggtctaa 540

agttctgaga tgttagcaca tacccttttc acaatttagg aagctttaag atcatttagt 600

atttttttat gttacaaaat ttggtacaat acacctcttt caggaaagtc ttagtagtaa 660

ctccaaatat tataattatt gtaaccagaa ttgtgacact tggagcagaa tgcatgcaca 720

caaaataaaa tcctgtcaaa aaatgacatc accattcccc cacaccaaat gtgtaattgg 780

taggaaatgc atttccagtc tggtacatgg cagtgtgaca aactcctact cactcgcttt 840

tcaagttggt gactgcagct gaaatgtttt tctgtgatgt atgccaccct tttacctatt 900

tgatttggaa gtgtagaatt cggattcatg tcatctccac agacctttcc tcttaggagt 960

gcctaagctg tcttactctg atggaggtat aatgtagcac gaaagacttc caaagaacca 1020

gtttctctct tgctgttcct cttaacaact ttcacgtcta tctaaacatt ctatgcagga 1080

gtcctactaa gaaattttgg tgtaatgcca ctttgatcag ttatttgttg tatgacttca 1140

ttcaaaaaca ctttcatcaa tagcatgggg attgtatcta tgaaagggaa gttggtgtcc 1200

tgcgttcctc acaaaattat ccaaaggata aaatgaaaag tatgtgagaa acctgcttta 1260

at 1262

<210> 121
<211> 562
<212> DNA
<213> Homo sapien

<400> 121
ggtaccaact tgagtgcctc ctaaaagtgt aaccttgggg gcggggatac agaaggatga 60

tgctgacaat ggaatttaaa aacaaacagc aacattttgt ggtgtctaca ggcgtggggg 120

tggaggagct gcagcgtcac catgggaaca aaagtctccc acgcatctca ggcccgagga 180

atctttaaag agggagagtg ggcatgggag gaggacttaa gctattagtc atattttatt 240

tcgaaaacta gatcttaagt aactgtagca aaatgttaac aattcttacc ttggaatacc 300

ggttacatgg gattcatgtt actctatttt ttcatcatgt gcaaatattt tcatattttg 360

acaattaaaa ctaaatagta gctttttata aaagtggcat atgcactgaa gtataatgtg 420

ctaatttggg attcgtttaa ataaaacagc tttcttacaa aaaaaaaaaa aaaaaaaaaa 480

aaaaggttgg gggaaacaag ggcaaaaggg gttcccgggg ggaaatggtt accgggtcga 540

aatttcacaa ttggagaaaa ac 562

<210> 122
<211> 695
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (13)..(13)
<223> a, c, g or t

<400> 122
ctggagcatg gtntgcagga gtgcaagact gcaagcctcc tccacggcca ccactccagg 60

cctggataaa gaattcgtgg catatttcag ggaacagaat gtcccctggg gcgaaagggg 120

atgaagtcat tctacttgta ccaacttgag tgcctcctaa aagtgtaacc ttgggggcgg 180

ggatacagaa ggatgatgct gacaatggaa tttaaaaaca aacagcaaca ttttgtggtg 240

tctacaggcg tgggggtgga ggagctgcag cgtcaccatg ggaacaaaag tctcccacgc 300

atctcaggcc cgaggaatct ttaaagaggg agagtgggca tgggaggagg acttaagcta 360

ttagtcatat tttatttcga aaactagatc ttaagtaact gtagcaaaat gttaacaatt 420

cttaccttgg aataccggtt acatgggatt atgttactct atttttcat catgtgaaat 480

attttatatt ttgacaatta aaactaaata gtagcttttt ataaaagtgg catatgcact 540

gaagtataat gtgctaattt gggattcgtt taaataaaac agctttctta gaataaaaaa 600

aaaaaaaaaa aaaaaaaggt tgggggaaac aagggcaaaa ggggttcccg gggggaaatg 660

gttaccgggt cgaaatttca caattggaga aaaac 695

<210> 123
<211> 386
<212> DNA
<213> Homo sapien

<400> 123
aacccctggc caggcccagc tgccacaccc tttctgggag aagcatggcc tacagaatga 60

agagggggac caggaacccc tgtgggagag gcttagacct gaagcagtgc ccactctggc 120

tcctcctgcc ttggctgact gggttcctgg accatgtgca tttcactggg ccatgggatc 180

tacatctcct tgcatcccca gctggtctga tccctgccag ggccccttcc ttcctgctca 240

tggtcttcag gtggcctgat catggaaagt aaggagttag gcattacctt ctgggagtga 300

accctgactc catcccccta ttgccaccct aaccaatcat gcaaacttct ccctccctgg 360

ggtaattcaa cagttaaaag aagctt 386

<210> 124
<211> 654
<212> DNA
<213> Homo sapien

<400> 124
atgataaacc acctcagccc ccaccaagcc gccgcacccg tagaccagac cccaaggacc 60

ctggccacca tgggccagag agcattacct tcatctctgg ctctgctgag ccggcccttg 120

agtcccccac ctgctgcctg ctctggcgac cctgggtgtg ggagtggtgc cgggctgcct 180

tctgcttccg ccgctgccgg gattgcctcc agcgctgtgg aggccgtgtg cggggatgca 240

gcccctgcct gtctactgag gactcccctg aggggactgc tgaagccaac tggtccaagg 300

agcacaatgg agtgcccccc agccctgatc gtgcagcccc ccgccggcgg gatggccagg 360

cgggctgcaa gtcaaccatg ggcagcagct tcagctaccc cgatgttaag ctcaaaggca 420

tccctgtgta tccctaccga gaggccacct ccccagcccc tgatgcggac tcctgctgca 480

aggagccact ggccgatccc ccacccagcg agcacagcct gcccagcacc tttgccagta 540

gtcctcgtgg ctccgaggag tactattctt tccatgagtc ggacctggac ctgccggaga 600

tgggcagtgg ctccatgtcg agccgagaaa ttgatgtgct catcttcaag aagc 654

<210> 125

<211> 684
<212> DNA
<213> Homo sapien

<400> 125
acatgcagat gtgcatgtta cagagataaa gtgatcgaga caaggactga ctgggtatag 60

aaggaagaca gactcctgtc ttcactccta aatgcagttc tttggaatca ccctactgtg 120

atgggcgtag tagggagcca tcagctagga agaaacgtgg gagatgtgaa ttccaagagt 180

tgcctggaca gggcaagtca tgttagcgtg ggtcacactt ccaagatatt taaagcaaat 240

acaaaacaga acagaggatt caaaccgcaa gtatgggaga tttaggccct gcagaggcag 300

accattcctt agtatctcac aaagcagagt aatactggag gcagagtagg gggtggttgg 360

agagcagtta gtaccaataa caatgaagtc tgtgtttgat ctgatcgata ctttccagtc 420

ccgaatcaaa gatatggaga agcagaagaa ggagggcatt gtttgcaaag aggacaaaaa 480

gcagtccctg tgagaacttc ctatccaggt tccggtggag gaggaggttg ctggtgatct 540

ctgtcctaac gatgaagact gggctattca caggcagctc tctgccctca gtggtcaggc 600

gtgcacattt ggtctgcgcc acataacatt ctgaagcttg ggtatcatgg tcatagtgtt 660

ccgtgtgaat gtatcgtcac atcc 684

<210> 126
<211> 2671
<212> DNA
<213> Homo sapien

<400> 126
ctgccgaaga gttcaaaaca gaagagcaag atgcctcagg gagtatagaa tttggtgtat 60

cttttcctga tagggaatca tcatctatgg aaacatccat cgaaccaaaa gcaactgaaa 120

cttctcacac agagggaatt actgccattg aggagagctg ggagtctatg tttaacgatg 180

atggtgactg cctggatcca cgtcttctac aagagttatc agggaatacc aagagcagag 240

agagcatcca ggaacctaga tctgattact acaatcatga agttcctgat attgacctca 300

gtgattgtga attcccacat gtcattgaaa tttatgactt tccccaagaa tttcgtactg 360

aagaccttct acgggttttc tgcagttatc aaaagaaagg atttgatatt aaatgggtgg 420

atgatacaca tgccctagga gtattctcca gtccaattac agctcgtgat gcgttgggta 480

ttaaacacac catggtgaag attcgtccct tgtcacaggc cacaagagca gccaaggcca 540

aagctagagc ttatgctgag ttcctccagc cagcaaagga gcgtcctgag acttcagcag 600

ccctagccag aaggttagtc atcagtgccc ttggggttcg aagtaagcag agcaaaaccg 660

aacgagaagc agagctcaag aaactgcaag aagccagaga gagaaagcgg ttggaagcca 720

agcaacggga agacatctgg gaaggcagag accagtctac agtttgaaca tcactcaatg 780
aaagggataa ttccatgaat cagaaaatgt ttccatagcc ttcagataag atgatccttc 840
cagagctcta tgtacatgca gatgtgcatg ttaaagagat aaagtgatcg agacaaggac 900
tgactgggta tagaaggaag acagactcct gtcttcactc ctaaatgcag ttctttggaa 960
tcaccctact gtggtgggcg tagtagggag ccatcagcta ggaagaaacg tgggagatgt 1020
gaattccaag agttgcctgg acagggcaag tcatgttagc gtgggtcaca cttccaagat 1080
atttaaagca aatacaaaac agaacagagg attcaaaccg caagtatggg agatttaggc 1140
cctgcagagg cagaccattc cttagtatct cacaaagcag agtaatactg gaggcagagt 1200
agggggtggt tggagagcag ttagtaccaa taacaatgaa gtctgtgttt gatctgatcg 1260
atactttcca gtcccgaatc aaagatatgg agaagcagaa gaaggagggc attgtttgca 1320
aagaggacaa aaagcagtcc ctgtgagaac ttcctatcca ggttccggtg gaggaggagg 1380
ttgctggtga tctctgttcc taacgatgaa gactgggcct attcacagca gctctctgcc 1440
ctcagtggtc aggcgtgcaa ttttggtctg cgccacataa ccattctgaa gcttttaggc 1500
gttggagagg aagttggggg agtgttagaa ctgttcccaa ttaatgggag ctctgttgtt 1560
gagcgagaag aaaaaaaaga tgaagaatga gaacgcagac aagttactta agagtgaaaa 1620
gcaaatgaag aagtctgaga aaaagagcaa gcaagagaaa gagaagagca agaagaaaaa 1680
aggaggtaaa acagaacagg atggctatca gaaacccacc aacaaacact tcacgcagag 1740
tcccaagaag tcagtggccg acctgctggg gtcctttgaa ggcaaacgaa gactccttct 1800
gatcactgct cccaaggctg agaacaatat gtatgtgcaa caacgtgatg aatatctgga 1860
aagtttctgc aagatggcta ccaggaaaat ctctgtgatc accatcttcg gccctgtcaa 1920
caacagcacc atgaaaatcg accactttca gctagataat gagaagccca tgcgagtggt 1980
ggatgatgaa gacttggtag accagcgtct catcagcgag ctgaggaaag agtacggaat 2040
gacctacaat gacttcttca tggtgctaac agatgtggat ctgagagtca agcaatacta 2100
tgaggtacca ataacaatga agtctgtgtt tgatctgatc gatactttcc agtcccgaat 2160
caaagatatg gagaagcaga agaaggaggg cattgtttgc aaagaggaca aaaagcagtc 2220
cctggagaac ttcctatcca ggttccggtg gaggaggagg ttgctggtga tctctgctcc 2280
taacgatgaa gactgggcct attcacagca gctctctgcc ctcagtggtc aggcgtgcac 2340
attggtctgg gcgcccttacc ttctgaagct taagcgtgcg cacggactgg gggcccgttc 2400
aactggcccc attaagggac cccgagataa cgagaaacgt acaccccatg gtgaaaaaca 2460

ccgcacaaat ccacggaccc ggagacaacc caggccaggc gcaaaaagca agaccacacg 2520

gatatcaccc aaggcagcga gaagggacca cacacacacc cgcacaacag gacacccaag 2580

cggcgccaca acagtcacga caccacaagg ccacgaagca acacacagaa acatacacag 2640

cagcacacgg ccatacaacc gcccacacag c 2671

<210> 127
<211> 420
<212> DNA
<213> Homo sapien

<400> 127
acgggccgca gtgttgatgg atgcggcgag gtaactctct ctcccttaag agttatgagt 60

tatcaagagg agacttctta aagacagcaa cgcaattctt gtaacttgtg taaatagccc 120

catctttcag agtgatacca tttctacatt tgataatgcc tgtattcctg taggatgtat 180

atagtttagg ggattttttt ttggttgggg ttttggtttt ttagaaggtc aatatgtctg 240

gttttattta tgtgcttgaa aaagatcatt tgaaaaaaat aaatacattt tcaaccacaa 300

aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa ggcgcggggg ggaacccggg gcccagagcg 360

ggccccgggg ggcgaattgg gttctcccgg cccacattcc cccaaaatat tggcacacag 420

<210> 128
<211> 2269
<212> DNA
<213> Homo sapien

<400> 128
taccgaggag ggaacaagct acatgctatt ttgtttgtag tattgtggaa cagtcttgtt 60

atggagtgcc agcttagagg ttgttgcaaa cttgtctaga agtgagagca tggttttttt 120

tagccctttg agagtctaca tctaatgaac attcttgctc acccataaat aacgtcaagc 180

ctcaatgtca ccgtcacgtt gggatactct ttctcatctg gcatcctaga caggacaagg 240

ttggttacct ttccttccat gaaccatgaa cctgtgacgg catcattcat cctgacttca 300

ccaagctccg cctgtgggtg aggccagagc tcccactggc aatttttaga agagccagag 360

gctccctgct tcctctagaa ataacagttc agggtgaagc atggagggtt tcagttccca 420

gacaatggaa ccatttagag acaacacagt tggacatttc cactttttcc ttgattcctg 480

gaagtccagt gggttctgca gctgaaaaag ccctgggtcc cagcagcaga gagacaggac 540

agaggggatg cttgggcggg gagggacggt aacctgcaga acagattcca tttttataga 600

acgagtacac gtttgctaaa acagtcctgc tttcccagac tggattccca ccacagggac 660

agtcggaact caggactagc tccagcgaca tctttcctcc gaattcaagc cttctatcac 720

aatgtcaaaa cagctattta taaagccatt ttcattgtac ttgataacag cacgagtccc 780

aaaactttta gaaataaaat aggacattgg cttgattgaa aagagggact ttttaaaaat 840

tgttctttcg tcagaagcct tttggatgac ttacaatagc tctgatgaag ataccacccc 900

agcgtcagtc caataggtca gtgagtttca acaggcatcc atccctccca tgaagggatt 960

ctggtgatgg gaagtttctg taatgacagg aaagcattga ccctcattga ttgtcaactt 1020

tggtattagc catgaaagac aggatgctca ttgggtgttc tgtagagtga ggaatgctgc 1080

ctattccctc ccagaacgtc tgacccaggg gtgtgtgttg aggagccctg gggaaatgg 1140

accaagtttt cccacagagc agtattaggc tgaagagcag gtgactggta ggccccagct 1200

cccatcattc cctcccaaag ccattttgtt cagttgctca tccacgctgg attccagaga 1260

gttttccaat ttgggaagcc atgagaaagg tttttaaatc ttgggaagat ggagagaggg 1320

acataggata gttgactcca acatgacagg aagaggctgg agattgggaa ttggccatca 1380

accaagcctg tagtagtaaa gccatggtcc cgcattggaa ttacttgggg aacttataca 1440

gttctgatac ccaggctctc ctagaccagt tcaaccaatt ctaggtgggg gactcaggca 1500

tcagtgtgtt tcgtagctcc ccgggtgttt tccctgtgca gccgagcttg ggaaactgcc 1560

atgctttttg gatgtcaagg cgctgttgga ggctgggtgt gacagcacag agccaggttg 1620

tcttgtggaa accacagcca cgggtttgcc actggctcag catggcctca ctgccagtcc 1680

cagcctggct gagggacaag atggtttctc ttgggagttc ctgagtggag caccccttcca 1740

ggctttttga aagccagctg atctgtggag ccttgttaag ggactcaata cggtgtttgg 1800

atattgatgt ttttccttga gactgtcttg tccatcaata aagatggagg atgtctcctc 1860

tttgaacccc gcttccccac cagtactctc tctcccttag agtttatgag ttattcaagg 1920

aggagacttc ttaaagacag caacgcaatt cttgtaactt gtgtaaatag ccccatcttt 1980

cagagtgata ccatttctac atttgataat gcctgtattc ctgtaggatg tatatagttt 2040

aggggatttt ttttttgttt ggttttgttt tttagaagtc aatatgtctg gttttattta 2100

ttgcttgaaa aagatcattt gaaaaaaata aatacatttt caaccaaaaa aaaaaaaaaa 2160

aaaaaaaaaa aaaaaaaaag gcgcgggggg gaacccgggg cccagagcgg gccccggggg 2220

gcgaattggg ttctcccggc ccacattccc ccaaaatatt ggcacacag 2269

<210> 129
<211> 750
<212> DNA
<213> Homo sapien

<400> 129
gccgcccggg caggtaccca agtttcagtt acacaggagg catgagattg atctagtgca 60

aaaaatgatg agtataataa ataataatgc actgtatatt ttgaaattgc taaaagtaga 120

tttaaaattg atttacatac atattttaca tatttataaa gcacatgcaa tatgttgtta 180

catgtataga atgtgcaacg atcgagtcag ggtatctgtg gtatccacca ctttgagcat 240

ttatcgattc tatatgtcag gaacatttca agttatctgt tctagcaagg aaatataaaa 300

tacatttata tgttgactat ggcctatcta catgttgcaa ctaaacacta gattttactt 360

cctttccaac tgtgggtttg tattcattta ccaccctctt ttcattccct ttctcaccca 420

cacactatgc cgggcctcag gcatatacta ttctactgtc tgtctctgta agcgattatc 480

agttttagct tccacatatg agagaatgca tgcaaagttc tgtctttcca tgcctggtct 540

tatttcactt aagcaaaatg acctccgcgt tccatccatg ttatttatat tacccaacta 600

gtgttcataa aactagtata tacaccacat agtataccac agaaacggac cactgcggat 660

aaacaggatt tctggtccac acttttgtcc catacgggac cgtggggcaa tctgattacg 720

cgcacagcaa gagcaaccca gtaagaaaca 750

<210> 130
<211> 738
<212> DNA
<213> Homo sapien

<400> 130
gcgtggtcgc ggccgaggta ctgtgaatta cggatgctct ttgaaggaaa gaaatatcga 60

ttctaatgtt cttcagaagt tctggcaggg ataagcagga catcgactgg aacgtatgct 120

aaatgaaagc agacaaattt ctattttctt acctgagcaa atattttatt gaaactgctt 180

atgtatgcca aaggagccca caacttcagc tacacaactt tttgtattga aagaactcat 240

actttttgta gcttttattt cacatttaat ttaaagtgac ttttagcact aaaatgccta 300

gaagatttta ctccagacct ataaggaaat gtttagtttt tatgaaaaat gacaagtcga 360

tggttaaact tctcatgtct ttggtgcttt ggccctaata gcactggaca acaccacgac 420

cacatggaaa catatttttg gaagcaaaac tttaatttta tataacgtat gctatggaga 480

gctaagacaa tttaaggact acttgttttc tattttttttt cttaataaaa tggaatccac 540

tgtgttgaag actcttgata ttcatgtgct tgtctaacca ttttttgttt tataattaga 600

ataaaatata gttgtgataa tggtcatcga atggattttg tttggaaagc tacatcttat 660

ttgtgaaatg ttttttaaaa tcagagtaac tatcaactga ttcagctttt tgttgttttg 720

ttcttggtat aatacttg 738

<210> 131
<211> 1875
<212> DNA
<213> Homo sapien

<400> 131
tggcaacgat ctggaccgct acaacccgct aagctccagc gccttgtgcg caacgcgctg 60

gcgcacgtgg tgccaaggag cgcgagctga gctggcgcac tcggagagtt tcgccgcctg 120

tgccgctacg gcaagcgcga gttcaagatc ggcggcgagc tgcgcatcgg caagcagccc 180

taccggctgc agattcagct gtcggcgcag cgcagccaca cgctcgagtt ccagagtcta 240

gaggacctga tcatgggaga agcgacgcaa cgacccagat cgggcgcgcg gcccgtgctg 300

caggagctcg ccacgcacct gcacccggcg gagccggagg agggcgacag caacgtggcg 360

cggactacgc cgcctcccgg gcgccccct gcgcccagct ccgaggagga ggacggagag 420

gcagtggcac actgatgggc gagctgagcg cagagctgcg aaggggaact gtttgcagta 480

gcagccgctg ctccctttct ccctctcttc ctccctcttt tgccactgtc tgggccccat 540

ctgggattcc tgggcccttt ggaaaagagt tggtgaaatg cgcagccggc tgtggacggg 600

ggaggaggaa ggggacagag ggagcaggaa taagactgta gaactgtttt gtactgtgaa 660

ttacggatgc tctttgaagg aaagaaatat cgattctaat gttcttcaga agttctggca 720

gggataagca ggacatcgac tggaacgtat gctaaatgaa agcagacaaa tttctatttt 780

cttacctgag caaatatttt gttgaaactg cttatgtatg tcaaaggagc ccacaacttc 840

agctacacaa ctttttgtat tgaaagaact catacttttt gtagctttta tttcacattt 900

aatttaaagt gacttttagc actaaaatgc ctagaagatt ttactccaga cctataagga 960

aatgtttagt ttttatgaaa aatgacaagt cgatggttaa acttctcatg tctttggtgc 1020

tttggcccta atagcactgg acaacaccac gaccacatgg aaacatattt ttggaagcaa 1080

aactttaatt ttatataacg tatgctatgg agagctaaga caatttaagg actacttgtt 1140

ttctattttt tttcttaata aaatggaatc cactgtgttg aagactcttg atatcatgtg 1200

cttgtctaac catttttttgt tttataaatt agaataaaat atagttgtga taatggtcat 1260

cgaatggatt tgtttggaaa gctacatctt atttgtgaaa tgttttttaa atcagagtaa 1320

ctatcaactg attcagcttt ttgttgtttt gttcttggct ataatacttg tgactcatga 1380

agaattatgt tgacaaacag gataaattcc acatgcattt tatttcccag tgagttgtat 1440

aaactttatt tttgttgaag gttgtatgtt aaatcaatgt tacattctta tatcacttct 1500

tgagaaggaa gttccgattt gaaattgtat catttccttc aaaatgaagg gcagtgctta 1560

gttaaataaa agattgatga tatcttttaa gccaaaaaaa aaaaaaaaaa aaaaaaaaaa 1620

aaaaaaacga accaaaccaa taaaaacaag aagcacacag accgaacacc acacacacaa 1680

gccaccagag ctcacataac gcgcgggcaa acatccacac ggccacacac agcaacccac 1740

tatgagagcc accccgcgga acaaaagacc ccacacacaa ccagagacaa gaaacctgcg 1800

agccacgccg tccacaccca caaccacgaa tagtcacctc agtaacaaaa caaacacaga 1860

cggaggcgcc gacaa 1875

<210> 132
<211> 828
<212> DNA
<213> Homo sapien

<400> 132
tggtcgcggc cgaggtacaa taggtctctt gaatttattc ctcctgtcta attgaaattt 60

gtatcccttg accaacatct tcccagtcac acccccatcc ctctggtaac catcattcta 120

ctctagttgt atgagttcaa tttttttaga ttccatttat aagtgattta attaatatct 180

ttatcctctt tccagataat tcaaggacct tagcatttta actctagtca actgtaatat 240

tacattccat cgtattgcag tattttagtc ttcttctatt aagccttcca aattggatat 300

tagcattatt gtggttgttt cacattagca ttattgtggt tgtttcagat agtcaatatt 360

gatgcagatt tacctgaata ttacccatga ttaccatcat tccttctttc tacttagatt 420

tccatcatcc ttcttcttga aatataattt ttaaaaggtc cattgaagaa gttctgttga 480

tggtaaatac agttttactt tctttgaaaa tatctttatt ttgcccacat cagttatttt 540

attgttcagt attaagaaaa cctaattcct gtgttttctt cccatcattg ttgatattga 600

gttgtgtgcc atcaggcaaa tgtcattact ttttagatat tctaaacctg ttgtttcttt 660

aagtaagtac attgtctccc ccttaatctg ttctccttcg taatgtttta ttatttgtct 720

cactattatg gattctggac aggtttcttc tgggtccttc tttcaggttg ctattctcta 780

ttcaggtgtg tttatctgct atttatcatc cctccagttt tttccttg 828

<210> 133
<211> 1023
<212> DNA
<213> Homo sapien

<400> 133
tggtcgcggc cgaggtacaa taggtctctt gaatttattc ctcctgtcta attgaaattt 60

gtatcccttg accaacatct tcccagtcac acccccatcc ctctggtaac catcattcta 120

ctctagttgt atgagttcaa ttttttaga ttccatttat aagtgattta attaatatct 180

ttatcctctt tccagataat tcaaggacct tagcatttta actctagtca actgtaatat 240

tacattccat cgtattgcag tattttagtc ttcttctatt aagccttcca aattggatat 300

tagcattatt gtggttgttt cacattagca ttattgtggt tgtttcagat agtcaatatt 360

gatgcagatt tacctgaata ttacccatgg attaccatgc attccttctt tctacttaga 420

tttccatcat ccttcttctt gaaatataat ttttaaaagg tccattgaag aagtgtctgt 480

tgatggtaaa tacagtttta ctttctgttg aaaatatctt tattttgccc acatcagtta 540

ttttattgtt cagtgattaa gaaaacctaa ttcctgtgtt ttcttcccat cattgttgat 600

attgagttgt gtgccatcag gcaaatgtca ttacttttta gatattctaa actgttgttt 660

gctttaagta agtacattgt gctcccctta atctgttctc ttcgtaatgt tttatttatt 720

tgtctcacta taatgaattc tggacaggtt tcttctggtc tttctttgca gtttgctaat 780

tctctattca gctgtatcta atctgctatt taattcatcc atcaagtatt ttttccttag 840

tattttgttt taataatttt atttactatt tctagatttt tttctaatca tcctggtctt 900

tgtcatagta tcttcttctt tatatacatt ttatttatgt atctgataac attaataact 960

taaacctttg taagttataa gtatgttttt agttttggtg ctgatttggt tcaaataaac 1020

ata 1023

<210> 134
<211> 757
<212> DNA
<213> Homo sapien

<400> 134
gagcggcgcc cgggcaggta ccttcgtgcc cctcagtagt tgttttagcc taatgtagag 60

tcaatctagg acttataatt attcatcatg attttgagta gattgtaatc atcaagaatt 120

tttcatagat cgtttacttc caattgaatt tagctcagaa gtgattgctt tctctttatt 180

tgagatagga gctctcgcac tgtcgccagg ctaggagtgc aagcggtcat gatcgtcggc 240

tcactagcaa cctctgcctc ccgggttgaa gcagatatac ccctgacctc aagcctcctg 300

cagtagctag ggactacagg tagttcatcg cttgtcctta gcttggaaac taggatgcac 360

aaacacatgg gttattatac tcgtacacgg agctggtcac acaacggaac tagactctct 420

ctccaaatgt gataccacac agacaacact cagaactacc ttcgagcctt acttaagatc 480

atcccttcac tgatctaaca aacttacaaa cattaataca accagatact gcgtctcgac 540

tattgcacgg caaatcaaaa tacaacaggt tctccactaa agaccaggtg gtgacatgtc 600

ctagagatca acagaacaat ctaatcctga ccctcacgcc aactatgatg acacgatggc 660

cgctggccca cacaggaagg ccgacacggg ccgcgctcaa agaccaccca tgtccggacc 720

tagcctaaaa aaaactcacg ccccgccgcc cctacct 757

<210> 135
<211> 1513
<212> DNA
<213> Homo sapien

<400> 135
gcgggagcct gggcggcgag ccgggtgtga gctgcctgaa aatgcactcg gatgccgccg 60

ctgtcaattt tcagctgaac tctcatctct caacactggc aaatattcat aagatctacc 120

acacccttaa taagctgaac ctaacagaag acattggcca agacgatcac caaacaggaa 180

gtctgcggtc ttgcagttct tcagactgct ttaataaagt gatgccacca aggaaaaaga 240

gaagacctgc ctctggagat gatttatctg ccaagaaaag tagacatgat agcatgtata 300

gaaaatatga ttcgactaga ataaagactg aagaagaagc cttttcaagt aaaaggtgct 360

tggaatggtt ctatgaatat gcaggaactg atgatgttgt aggccctgaa ggcatggaga 420

aattttgtga agacattggt gttgaaccag aaaacgtgag tcaaacttac tgagttgggt 480

gaatcagttg gttgtttttc atacttaaat ctttgttctt tagcaaataa atagaataat 540

taaaaagtag tggtatgtta gtttttatga agcagtctaa gaaataagtt ctaattctag 600

tttgacttat aagcagattc tccattcttg taagtgatat ggtgtaacta cagttatttt 660

ttctctcatt taatttcttg tatgtaaaag gtacagtaag ccagatgctt acaaaatggt 720

gtggccacat gtgcctacaa tgacggatca actggaggcc acattgtacg ctgtgtacct 780

tcgtgcccct cagtagttgt tttagcctaa tgtagagtca atctaggact tataattatt 840

catcatgatt ttgagtagat tgtaatcatc aagaattttt catagatcgt ttacttccaa 900

ttgaatttag ctcagaagtg attgcttttt tttttttgag ataggagctc tcgcactgtc 960

gccaggctag gagtgcaagc ggtcatgatc gtcggctcac tagcaacctc tgcctcccgg 1020

gttgaagcag atatacccct gacctcaagc ctcctgcagt agctagggac tacaggtagt 1080

tcatcgcttg tccttagctt ggaaactagg atgcacaaac acatgggtta ttatactcgt 1140

acacggagct ggtcacacaa cggaactaga ctctctctcc aaatgtgata ccacacagac 1200

aacactcaga actaccttcg agccttactt aagatcatcc cttcactgat ctaacaaact 1260

tacaaacatt aatacaacca gatactgcgt ctcgactatt gcacggcaaa tcaaaataca 1320

acaggttctc cactaaagac caggtggtga catgtcctag agatcaacag aacaatctaa 1380

tcctgaccct cacgccaact atgatgacac gatggccgct ggcccacaca ggaaggccga 1440

cacgggccgc gctcaaagac cacccatgtc cggacctagc ctaaaaaaaa ctcacgcccc 1500

gccgccccta cct 1513

<210> 136
<211> 738
<212> DNA
<213> Homo sapien

<400> 136
gcgtggtcgc ggcgaggtac caaccccagc acaccccaac agcctttcct cggcccсctc 60

ctcaggcctc ctaattactc tttctcagcc tggagtgtgg ggccgttacc gtcctcttcc 120

cccttctcct tccatactgc acttaacctt gctggaagat ttaatgatgg agatttaggg 180

caactgtggc tgcttgggac ccttccctgg gaccaaagga acttaaaacc caatacctga 240

cactggaatg aaatccaagt ttttaaatat cacctttcaa tcactcacag atctcacatc 300

tatcttaaaa tactcagcct cactccttaa ctgagtgctt gcctgagagg gagaaaagtt 360

ccattttaaa aacgtattca ctttactgat tactgtgcaa tttgaattaa gtcacgattc 420

tttagtcatg gaggtcgaga atctcagatt caaattgtca gagaccatga tttagaagtc 480

taccaaacac ccagtttcct tccactgttt tagggtaaca ggaaaacatg agattggggt 540

ggtgtccgct attaaatgga accacacatc atgaaattca attctcatgt taagacattc 600

tgtattgtgg gatgtcaaaa gtatctccca aaactttcgt ttgacctgtc agagtgggga 660

tggttactcc ctatacttca gtttgtttca caagcttggc gtaaccaggc atagtgttcc 720

gtgtgaatgt tcgtccac 738

<210> 137
<211> 1350
<212> DNA
<213> Homo sapien

<400> 137
atggttatgg agaagcccag tccgctgctt gtagggcggg agtttgtgag gcaatattat 60

actttgctga ataaagctcc ggaatattta cacaggtttt atggcaggaa ttcttcctat 120

gttcatggtg gagtagatgc tagtggaaag ccccaggaag ctgtttatgg ccaaaatgat 180

atacaccaca aagtattatc tctgaacttc agtgaatgtc atactaaaat tcgtcatgtg 240

gatgctcatg caacccttgag tgatggagta gttgtccagg tcatgggttt gctgtctaac 300

agtggacaac cagaaagaaa gtttatgcaa acctttgttc tggctcctga aggatctgtt 360

ccaaataaat tttatgttca caatgatatg tttcgttatg aagatgaagt gtttggtgat 420

tctgagcctg aacttgatga agaatcagaa gatgaagtag aagaggaaca agaagaaaga 480
caaccatctc ctgaacctgt gcaagaaaat gctaacagtg gttactatga agctcacccct 540
gtgactaatg gcatagagga gcctttggaa gaatcctctc atgaacctga acctgagcca 600
gaatctgaaa caaagactga agagctgaaa ccacaagtgg aggagaagaa cttagaagaa 660
ctagaggaga aatctactac tcctcctccg gcagaacctg tttctctgcc acaagaacca 720
ccaaagccaa gagtcgaagc taaaccagaa gttcaatctc agccacctcg tgtgcgtgaa 780
caacgaccta gagaacgacc tggttttcct cctagaggac caagaccagg cagaggagat 840
atggaacaga atgactctga caaccgtaga ataattcgct atccagatag tcatcaactt 900
tttgttggta acttgccaca tgatattgat gaaaatgagc taaaggaatt cttcatgagt 960
tttggaaacg ttgtggaact tcgcatcaat accaagggtg ttgggggaaa gcttccaaat 1020
tttggttttg tggtttttga tgactctgaa ccagttcaga gaatcttaat tgcaaaaccg 1080
attatgtttc gaggggaagt acgtttaaat gtggaagaga aaaaaacaag agctgcaaga 1140
gagcgagaaa ccagaggtgg tggtgatgat cgcagggata ttaggcgcaa tgatcgaggt 1200
cccggtggtc cacgtggaat tgtgggtggt ggaatgatgc gtgatcgtga tggaagagga 1260
cctcctccaa ggggtggcat ggcacagaaa cttggctctg gaagaggaac cgggcaaatg 1320
gagggccgct tcacaggaca gcgtcgctga 1350

<210> 138
<211> 569
<212> DNA
<213> Homo sapien

<220>
<221> misc_feature
<222> (509)..(509)
<223> a, c, g or t

<400> 138
cgcccgggca ggtcgcccat gtgctgtgat gtcagtgagc gggcggagtt caggctggtc 60
agtgccaggt gctccttctc ccacccgaga acagtggcca ggttgctcct caggcaccct 120
gggcaactgc cccttccctt ccagtggggc ctgacctggc taccgagctt ggcagctaat 180
aggcgggccc ctcagcattc acgctcctga gctgctttat caaactagga ttgttccccc 240
aggtctaaga aaaccatcca ttcactgcaa agttagttat tactgcggat gggctaggag 300
ttagaggaag agagtgactc aaatcacaac acctcctgga cgaagctgga agcggattaa 360
aataccgggc ctaatttcag aacaacaaaa aaaaaagaaa aaaaaaaaaa agcgcgggcc 420

ggaacccagg ggccaaaagg gtgggtcccg ggggggggaaa tctggttacc gcggcccaaa 480

attccccaaa aaatttgggg gggccaaang caccgcgctc tctgcccccc ccacgcccgc 540

cccccccccc acaacccatc gccgccccg 569

<210> 139
<211> 739
<212> DNA
<213> Homo sapien

<400> 139
tatatcacta taggggactg ggtcctctag atgctgctcg agcggccgca gtgtgatgga 60

tccgggcagg tactgcctgg ttttacaaga attaatgcag tttcacagtg aagcatgtaa 120

gatattgaat tttagagaca atagaccaga tacctttcta atctcatttt attcattaat 180

gtcaaataat accattttta aaaatatggt gcttatttgt ctagcaagta acctatagaa 240

aagtattatt ttatacaaaa agatgattag gtcacataaa ggaattggaa tcttaagttt 300

aaaatacact tctgttttta gccagaaggg agaaacgatg gttggattta tgccattttt 360

caattaaaaa ccatgtggta ctacttgaag cagtttctga gtaaatggag gtgtttaaag 420

atttgtatta ttctctccca atgactagat agtagtattt tacaatggag acttaaaagt 480

tttttgtgtt ttattctttc gcttttctat gccctcaatc caaagaacac cagaaataca 540

cttgtagtcg gaaaacttgg gtttatcact cgcatcaagg aatgacacac accatgggcc 600

actctggagc ctctcaataa aaggatgttt caaaggaaca acaacaaaaa aaaaaaaaaa 660

aaaacgttgg gggaaacaca gggcacaaag tgtcccgggg gaaattgttt tccgccacaa 720

tccaaaattc acaaaaaacc 739

<210> 140
<211> 1131
<212> DNA
<213> Homo sapien

<400> 140
aagttgatag tatatccacc acctccagct aagggaggca tctctgttac caatgaggac 60

ctgcactgtc taaatgaagg agaattttta aatgatgtta ttatagactt ttatttgaaa 120

tacttggtgc ttgaaaaact gaagaaggaa gacgctgacc gaattcatat attcagttct 180

tttttctata aacgccttaa tcagagagag aggagaaatc atgaaacaac taatctgtca 240

atacagcaaa aacggcatgg gagagtaaaa acatggaccc ggcacgtaga tatttttgag 300

aaggatttta tttttgtacc ccttaatgaa gcgtgagtaa gaatttcctt taaaggaaaa 360

tctttaaatc atgtaaatga tgacaatttt taaataatga gtatgaggtg aagaattcat 420

tttaaaacat ctttctgaaa tctcttgtgt atattcatat ttgtactgcc tgttttacaa 480

gaattaatgc agtttcacag tgaagcatgt aagatattga attttagaga caatagacca 540

gatacctttc taatctcatt ttattcatta atgtcaaata ataccatttt taaaaatatg 600

gtgcttattt gtctagcaag taacctatag aaaagtatta ttttatacaa aaagatgatt 660

aggtcacata aaggaattgg aatcttaagt ttaaaataca cttctgtttt tagccagaag 720

ggagaaacga tggttggatt tatgccattt ttcaattaaa aaccatgtgg tactacttga 780

agcagtttct gagtaaatgg aggtgtttaa agatttgtat tattctctcc caatgactag 840

atagtagtat tttacaatgg agacttaaaa gtttttttgtg ttttattctt tcgcttttct 900

atgccctcaa tccaaagaac accagaaata cacttgtagt cggaaaactt gggtttatca 960

cttgcatcaa ggaatgacac acaccatggg ccactctgga gcctctcaat aaaaggatgt 1020

ttcaaaggaa caacaacaaa aaaaaaaaaa aaaaaacgtt gggggaaaca cagggcacaa 1080

agtgtcccgg gggaaattgt tttccgccac aatccaaaat tcacaaaaac c 1131

<210> 141
<211> 887
<212> DNA
<213> Homo sapien

<400> 141
gcgtggccgc ggccgaggta cactgaatta ttcacagtaa tcgcttggtt ggggaaaggg 60

ttagtaaatg ccaaaggaaa tacccacaga aatctcctac acagcttaga tgttgtgctg 120

gcatttaagg cccatgagtg atggtccatt ctgcagcttt tcatgccatg cctttccttt 180

gtgtgggggt ccacagatca gagtctgtct gtggcatcga cttccttatg tcctcattgt 240

tcccacccat tgctgggatg tccacgttgg acttctcaaa agtggcccaa gaatctaagt 300

gcaaaatctg tttggatttt tacaattttt tcctaatctt ttacagtctt ggtcattcct 360

atttcaactg caattttttt caatgacttg cctggtgtga atattttttt aaagcatcca 420

gtattaaaca aaaaaattta aacagctaaa aaaaaaaaac aaaaaacaaa cggctgggcg 480

aaaccagggc tcaataccgg ctccccgtgg tgctgaacac tggtatactc cgcggttcac 540

caattcccaa ccacaacata cgggcgagac aaggctgcac gcaacccggc acgcgcatgt 600

cgcaggacac gtcacggagc caagaacggg cagcaggacc acagagaacc agacgcaggc 660

cgcgcacgtg gagcggaggg gtagaaccga cagccgccgc gccgtgggca gcggccatgg 720

cgcacacggg ccgacacgga agcggagccg cagcgacagc gagcagcacg cggggcgacg 780

gcgcggcgag gaggggagcg gcgcggggaa cggacgctgc agagaggcgg agggcggcga 840

gccgcggcgc ggccgagccg aaggcgaccg caagcggcgg cggcggc 887

<210> 142
<211> 2086
<212> DNA
<213> Homo sapien

<400> 142
cgagccaaga attcggcacg aaaaacaaat acttcctgat cgatcccttg tcttgtttag 60

tatgcttcct gaccattttt taccctaaca tttgtgttct tttcccgaga aggaaaatca 120

acttctatcc tatctctacc cagcagaggc ccctgcccca ctttacacac aaaaccatct 180

aactttttga tattctaaat gggggaaacc cctattttat aaccctcggt tacttttaat 240

ctttagatga ggaactagag gagccactat gttcctctca gcaccatgat ttatgcctta 300

gctaaggcct tcacttgggg aaggggaaga aggttgtttt caagcctgtg gcctcctgtc 360

actccccacc cctggaaggc ccttcacttt tgggtgatgc ctagaggcct catggacagc 420

agtcccttct gacacccagt gagatatcat ctgggagggt cgcagccctc agttcccctc 480

atggctctct ctttcacttc cctccatgac accacctcat cgagttgaag atgttattga 540

tgagtgcagt gggtgtatag tgtcctccca aaattcatgt ccacccagaa attcagaatg 600

caaccttatc tggaaataga atctttgcaa atgtgattag ttaagatgaa atcatactga 660

gttaggatga acctgaaatc caatcactgg tgtccttgta agaggaaagg tcacaaagag 720

acagaggaga tacacagagg agcccatgta atgatgggta cggagactga cgtggcacaa 780

ctataagcca aggaatgcca gggaaggcca gctagcagaa gctagggaaa aacacagagg 840

gattctcccc tggagccttt ggagggagtg tggccctgct gacaccttgg ttctggactt 900

ctggccccca gaactgtgag aaaataaatt tctgtggttt aagccacaca gtttgtggtg 960

ctctgacttc gtgagctttt ctgcccatct gacagcgcct gcctgccttc ctccctgccc 1020

accgtcctcc cgccccgtcc cagaccctcc tcgctcctca tcccactcca ctcctgtgag 1080

tgctcctcca caccatggct gcaatcccca ccttaagctg gggactccca aaccccgact 1140

tccccacagg gctcaggagg cctttctcca gccagcctca catttggact catgcttctc 1200

ccccatgcca ccctcagcta cgctgaatta ttcacagtaa tcgcttggtt ggggaaaagg 1260

ttagtaaatg ccaaaggaaa tacccacaga aatctcctac acagcttaga tgttgtgctg 1320

gcatttaagg cccatgagtg atggtccatt ctgcagcttt tcatgccatg cctttccttt 1380

gtgtgggggt ccacagatca gagtctgtct gtggcatcga cttccttatg tcctcattgt 1440

tcccacccat tgctgggatg tccacgttgg acttctcaaa agtggccaag aatctaagtg 1500

caaaatctgt ttggattttt acaatttttt cctaatcttt tacagtcttg gtcattccta 1560

tttcaactgc aatttttttc aatgacttgc ctggtgtgaa tattttttta aagcatccag 1620

tattaaacaa aaaaatttaa acagctaaaa aaaaaaaaca aaaaacaaac ggctgggcga 1680

aaccagggct caataccggc tccccgtggt gctgaacact ggtatactcc gcggttcacc 1740

aattcccaac cacaacatac gggcgagaca aggctgcacg caacccggca cgcgcatgtc 1800

gcaggacacg tcacggagcc aagaacgggc agcaggacca cagagaacca gacgcaggcc 1860

gcgcacgtgg agcggagggg tagaaccgac agccgccgcg ccgtgggcag cggccatggc 1920

gcacacgggc cgacacggaa gcggagccgc agcgacagcg agcagcacgc ggggcgacgg 1980

cgcggcgagg aggggagcgg cgcggggaac ggacgctgca gagaggcgga gggcggcgag 2040

ccgcggcgcg gccgagccga aggcgaccgc aagcggcggc ggcggc 2086

<210> 143
<211> 676
<212> DNA
<213> Homo sapien

<400> 143
gccgccgggc aggtactaaa taaaatgcaa aacatgtcac atcactcttc ttcatgggtt 60

catgtcctct gtgggtcagg tcttccacat gtagagtaga ggtagggtat gttcacacct 120

tcaatgacaa cctacacatt tctgctccaa caggtccaaa attgttccta ggtttcaaag 180

ttgttgtttg tttgtttttt tcctttttct tttttttttt tttttttgga gaagtggagt 240

ttggctctgg ttggccccgg tgtggagtgt gcaaggggcg gtgatctgcg gttcaccaac 300

aaacctcgtg gtcctccgcg gtttacaagg gcgattattc cgtggcctac aggcctcgcg 360

agtatagccg tgggatataa tagggcagtg gcgcacacca gtgcccgagc ttaatttgtg 420

ggtattttaa ggtagaagaa gcggggttct ctccccccctt tgtgtgggtc tcgagggcgt 480

ggactctggg aggcctcgcg tggaaccctc gaggggtgat ctcacacctg tgcgcttggg 540

ggccttccca caaaaggtgg gcctgggggg atttaccagg gcgtggcaga agcccaaact 600

atgtgggccg gggcgcacac aggggggttt cccaaaaggg tttttttaac cggtattaaa 660

aagagggttt cgctag 676

<210> 144
<211> 1260
<212> DNA
<213> Homo sapien

<400> 144
taaacataca cacatcaaaa ataactcagc cacatgcaac aatacagaga atcttaaaga 60

catagtatga gcaaaataat cagtacacac aaaattccac ttatacaaag ctcaaaaaca 120

aaattaagca atatttttta gaaatgcact tataaatgat gactgaccca ctatcaagga 180

aagtatttaa cattgctctg aaagttctgg aaattcttga ttttcctttc tcaatttcta 240

cacccatcac cacgcccagt cttccccaac tcactaaaca gcaccgtcat ccatttagca 300

tttcaagcca gtgagaagtc atccttaatt ctgctttttc attaatttcc ctacttctaa 360

tctattacgt gtcttattag atctaagatc aatatatttc ctgaatatgt ctatttatgt 420

ccatttccaa cactaccact gaagtctaag ccattgtcac ctttctttct ggattactgc 480

aatagcctca cagcttccac tcttgaccac atacactcca ttctgcactc agccctcata 540

gtgatcatta taaaggataa aatggtgtgg ccagttagct cagttggtta gatcatggta 600

ctaataaaat gcaaaacatg tcacatcact cttcttcatg ggttcatgtc ctctgtgggt 660

caggtcttcc acatgtagag tagaggtagg gtatgttcac accttcaatg acaactacac 720

atttctgctc caacaggtcc aaaattgttc ctaggtttca aagttgttgt ttgtttgttt 780

ttttcctttt tctttttttt tttttttttt ggagaagtgg agttttggct ctggttggcc 840

caggcgtgga gtgtgcaagt ggcggtgatc tgcggttcac caacaaacct cgtggtcctc 900

cgcggtttac aagggcgatt attccgtggc ctacaggcct cgcgagtata gccgtgggat 960

ataatagggc agtggcgcac accagtgccc gagcttaatt tgtgggtatt ttaaggtaga 1020

agaagcgggg ttctctcccc cctttgtgtg ggtctcgagg gcgtggactc tgggaggcct 1080

cgcgtggaac cctcgagggg tgatctcaca cctgtgcgct tgggggcctt cccacaaaag 1140

gtgggcctgg ggggatttac cagggcgtgg cagaagccca aactatgtgg gccggggcgc 1200

acacaggggg gtttcccaaa agggtttttt taaccggtat taaaaagagg gtttcgctag 1260

<210> 145
<211> 433
<212> DNA
<213> Homo sapien

<400> 145
cggccgccgg gcaggtactg gtggttggtt tcattagtgg atcacacaca gggttgtact 60

tggcttgtaa aatggtgcct cggatagggt gagtttggat aagtatgtat gtatgtatga 120

gttatagcaa aattaagtag attgaatcaa gtccatgcaa aagcagtaaa acagttatta 180

attgttaatt ttttaaaaat taaaacgtta ataaaacagt ttgtaatgtt ttgctagtgt 240

cttttataaa atgatgtaag ttacagtgga agtcttcaca ggacttgtgt ctttcctgga 300

actattgaaa tgtaatttag gatgatttga tcttccatct caagttgtca acatggctgt 360

gtcattctgg cttacatatg ttttatttaa caaaattcta gtcaagggat aaggccttaa 420

tgaagacaag ctt 433

<210> 146
<211> 1791
<212> DNA
<213> Homo sapien

<400> 146
ggaatgaaca aacaaacaaa aatccttgct ctcctggtgc ttacatttta gttgggagag 60

ggacaaacaa gataagggaa atacatacct tagttaagaa caagtgccac agaggaaaag 120

ccaggctgag gcagtgggtg tgaacatttt atacagggat gtccagaatc agggctttga 180

agaaagccct gaaggcagcg tgtaccgagc aggaatgccc tgtggaggct gagcatttag 240

gaagtgggaa cagccggtgc ggaggtcctg gagggtgagg ggtgtcaaga aggccagcat 300

ggctggagca gaaagcaggg cggggaggtg ggggaccagc tcacaggtgc ctagagccag 360

aatgagaagg gcttcttggc tggattacag gcgtgagcca ctggaacctg gccttgtttt 420

gctttatttt ttctcttaca tgaagtaaag cgctttggtc aaacacacaa aaatactgcc 480

ttgtactggt ggttggtttc attagtggat cacacacagt gttctacttg gcttgtaaaa 540

tggtgccttg gatagggtga gtttggataa gtatgtatgt atgtatgagt tatagcaaaa 600

ttaagtagat tgaatcaagt ccatgcaaaa gcaataaaac agttttaatt ttttaatttt 660

ttaaaaatta aaactttaat aaaacagttt ttaatttttt gctaggttct tttaaaaaat 720

gatgtaactt acatggaagt cttcacagga ctttttttctt tcctggaact attgaaatgt 780

aatttaggat gatttgatct tccatctcaa gttgtcaaca tggctgtgtc attctggctt 840

acatatgttt tatttaacaa aattctagtc aagggataag ggcataatga agacaagctt 900

cagttatgaa agtacaaact atttgtgtga ttaattttta aaaatgacat taagaagccc 960

attgtaaaat aatatttgca gtcaaatggt tttttcttgct gtaagtcctg ttgtagctat 1020

gtttagggta gtggttctca tctaccttgg agtgcataag acttacctag caggcttgtt 1080

taaaaagttc agattcctag ctttgtaccc agggattgcc tcaggtggta tgggctgtgg 1140

tcctggagtc atcactttta taaatagtgg ttcagagacc acagagagag actgcttcat 1200

cgaatgggaa gtaccaagga gaaagtacaa ttcagtattg tctggaggca agtggacact 1260

ttgtacctga ggtttagaat aggtggtctc ttgccagtac aatccccagg cgttttctgt 1320

gttcagaagt agtaagaatg cctttaattc agaggattat ctaagctctt taaagctgtt 1380

tttctccatt tgtcatagtg ccttctctga aaaatgaatg tacaggtatc ctattttcta 1440

atgtaattag gattttttaa aagcaatttt gatagttttt cttttaaaaa gtaaaattca 1500

gcactgtgac ttgaaccccc aaatctttca catacaggtg aaacattaag ccacaaataa 1560

atatgacaga aagaagaaaa gatcctattc ctgtcattag ggactagtac ccattaactt 1620

gaaccgactc ggcaaggttg caacatttct tggcacatcg tgcacacact atgttttgac 1680

acgaggactt cccacttata aacaccggac cggggaatat ttcacatcgt ttaagtaatg 1740

caccccgggc aaaaaggaga aaccctcatt caaaaaatct atcgccgtct a 1791

<210> 147
<211> 349
<212> DNA
<213> Homo sapien

<400> 147
ggaatgatcg atactatagg ggcatggttc atctaatgca tgctcgagcg cgcgccagtt 60

gtgatggatg cgtggtcgcg gccgaggtcc acgtttagct gagtataatt ttaaatagcc 120

ctatgtgaca agtggctact ttattggaca gtgtagatct aagattaatt cctcaactgt 180

tttgcactca acaaagacat acctctgagt tggcaaccag cagggtggat aacgggccag 240

tggtgataaa atcaaagaat aggtaatgaa acaatcatcc agttaacaat cagcaaggtt 300

cttcagagcc taattaatgt ttaattctaa ataaattgca acaattaag 349

<210> 148
<211> 848
<212> DNA
<213> Homo sapien

<400> 148
agctgggatt acagacgccc accaccacac ccagctaatt tttgtatttt tagtagagat 60

ggggtttcac catgttggcc gggctggtct tgaactccta acctcgtgat cctcctgcct 120

cagcctccca aagtgcaggg attacaggtg tgagccactg cgcccgacat cccatttaac 180

tttctgtctc tgtgactctg atgactctag gaacctcata taagtggaat aatataggat 240

ttattctttt ttaaaaaatt tattttgaga tggagtctca ctctgtcact caggctggag 300

tgcagtgact cgatctcggc tcactgcaac ctccgccttc ctggcttaag caatttttgt 360

gcctcagcct cccaagtatc tgagattaca ggcgtgtgcc accacaccca gctatttttt 420

atttttttatt tttagtagaa gatggggttt cgccatgttg gccggactgg tctggaactc 480

ctggcctcaa gtggtcctcc cacctcggcc tctcaaagtg ctgggattac aggcgtgagc 540

caccacgttt agctgagtat aattttaaat agccctatgt gacaagtggc tactttattg 600

gacagtgtag atctaagatt aattcctcaa ctgttttgca ctcaacaaag acatacctct 660

gagttggcaa ccagcagggt ggataacggg ccagtggtga taaaatcaaa gaataggtaa 720

tgaaacaatc atccagttaa caatcagcaa ggttcttcag agcctaatta atgtttaatt 780

ctaaataaat tgcaacaatt aagaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 840

actcggtc 848

<210> 149
<211> 414
<212> DNA
<213> Homo sapien

<400> 149
cagtggtacg cgcgacgcag gtaccacagc tcccagtgcc cattacctct atcatggatg 60

ctgggtgact ttgggaagtc accacctctt cccaagcctg tttcccatat cacagatgtg 120

gggccatggc ctcgatgatg gtctccacag gtctttccac ctctgtgagt ccaagtcagg 180

tcaatcagca aggacccaat ctctgaccct gggtcagctc ctcagaacca acccccagca 240

tctctaaagc aaaagcctca cctcaagggc tgctcagaag agagcacctt cagcatgagt 300

tgttgctgga aggatctaat aagctgtgtt tcctgggaag tggtgcttta cttagccctg 360

tggacaactt ctctatgcat ctgtgtgagc agatgatcat tgtattacct ttta 414

<210> 150
<211> 2088
<212> DNA
<213> Homo sapien

<400> 150
ggtggcagtg atacatgttg gcaggctggt cttgatcctg actcaagtga tccgccgcct 60

caacctccca aggtgctggg attacaggtg tgagccacca cacttgtgca gtatatcctc 120

agtatgaaga tatttttatc ttcctgtgtt ctctggcttc agagtttcac ctgcccacac 180

agggtccgtt gctggcaact ggacttcccc ataagccttg ggtatcctgt gatgggctgt 240

gtctccctga agattgtctg gcttgcccac ttctccgtgc atgactctgg gtgtgagtct 300

gtctaggaac aggagggaaa gttggactca gacagaaatc agatgcttcc atgtattcag 360

ggcgcgcatt gtgagcagtg gagtatgagc cttgagggcc tcatggttgc agggcaggct 420

tccctgcaga tgggtggcag cccctggtag aatgctggat ttctctggaa tctagaagtg 480

ccatatttta gtggaaaggc atcagggctg tttgacagtg tgcgtctttc caatcccatg 540

ttcctccatt cgtgtgtctg ttataaaact gagtgaaggc tgctatgacc tgtgttcact 600

ctggttacag ggaggtgcaa accattctgt ctcccagcct ttcttctctc tttgtgtgct 660

cccagcactt ccttcttttc taacatggcc tggagagagt ctctctctcc ttgtctctgt 720

ctcttaataa tagtttttaa cgtggacatc tcttccttgg tacagtggtt tttaaatcct 780

gagaagaacc aagtcaggtt ttttaaagca gactaaaagc atgaaattgc tttcagaaga 840

atgtatatca tcgggaaaag tttgggggca gagtggggga atcaggcttt attcaaaaga 900

aacagttgaa aacatggact ttttctaccc aatgcccatt tcacgactcc tctgagacta 960

attgggaaac ggggaaattc ttggaatttt ttttttaaga aacttttttg tgtttttttt 1020

aattttaggt cacttattag tgaaacctca ttttagatct gacattggta gatagatgga 1080

tttaggcaaa tatgatgcgt ttgtggggaa tccacgtggt tgacgttaga acctcccttc 1140

tgcagactgt tgcctgtcat ctaagcgaat tggaaatgct gagcttccat aagtcagctg 1200

agttttaaag gtaaacgtta tggctgaagt agtaaagcac ctgaccacaa aacctcttgt 1260

aaaaacagcc ctgagtaggt atttccaggg ctccacaaag ttgcttatgg gaatcctgag 1320

ctgcttttca ccatctcaag aagcctaaga agttatatat ttaatcaggt agacaaaaca 1380

gttcaaagca taaggtccat ggtggtggaa aatggatgca agtgattcta agtttgtgga 1440

tttgtggata gcagagggat cgggacctct tggaggaacc ctgggtacca agctcccagg 1500

cccttcctct atcatggatg ctgggtgact ttgggaagtc accacctctt cccaagcctg 1560

tttcccatat cacagatgtg gggccatggc ctcgatgatg gtctccacag gtctttccac 1620

ctctgtgagt ccaagtcagg tcaatcagca aggacccatc tctgccctgg gtcagctcct 1680

cagaaccaac ccccagcatc tctaaagcaa aagcctcacc tcaagggctg ctcagaagag 1740

agcaccttca gcatgagttg ttgctggaag atctaataag ctgtgtttcc tgggaagtgg 1800

tgctttactt agccctgtgg acaacttctc tatgcatctg tgtgagcaga tgatcattgt 1860

attacctttt atcggtagta agcttggaaa aataatttaa gaatacaatg gagaaatgta 1920

aataagtatc tatgtaaatt tgtttaaaat aaactgaatg tatttaatgg tccatttata 1980

tgttctttta tgtaacatgt agtttaataa agttcctgtt tatgagagtc atgtttcatc 2040

tcagcttctt ccaaaaaaaa aaaaaaaaaa agatctttaa ttaagcgg 2088

<210> 151
<211> 509
<212> DNA
<213> Homo sapien

<400> 151
cggactcccc ccgcggacgc gctggcttcg cgtatcggtt tacttccttt ataaaaattt 60

ttataactta tgtggaaatg ggatctcact atgttgctca gacttgtctt gaactcctgt 120

tatagcacca cctcttacaa gtgttgggca gctccgcttc tctcacttgt ctggaattct 180

aaggcacttc cctgaagtgc tcatcctgag ctaatatggg atagggctgg agagagaaca 240

gaggtggata gcatgccaga atgaggtggg aaggtgggga tcagcagcct ttgggaagga 300

aagaagtatg agtcccaggg tattaacaag gtggaggggc aataaaattt attacatatt 360

gggattcata ctaaatgagt agattttagc ttctcttgcc acaataaaca aaaaaaatgc 420

cacaacacca cacaaaaaaa aaaggtgcgg ggaaccaggg ccaaacgtcc ccgggtgaat 480

gtttcccgcc atcaaattaa aacacacag 509

<210> 152
<211> 560
<212> DNA
<213> Homo sapien

<400> 152
ccagcctggg taacagatgt gagaccctgt ctgttaagag aatcagaaaa gagagagaaa 60

gctagactta gctccaagtc tggagctttt ggggttttct tcctttataa aaattttta 120

acttatgtgg aaatgggatc tcactatgtt gctcagactt gtcttgaact cctggttagc 180

accacctctt acaagtgttg ggcagctccg cttctctcac ttgtctggaa ttctaaggca 240

cttccctgaa gtgctcatcc tgagctaata tgggataggg ctggagagag aacagaggtg 300

gatagctgcc agaatgaggt gggaaggtgg ggatcagcag cctttgggaa ggaaagaagt 360

atgagtccca gggtattaaa aggtggaggg gcaataaaat ttattacata tgggattcat 420

actaaatgag tagattttag cttctcttgc cacaataaac aaaaaaaatg ccacaacacc 480

acacaaaaaa aaaaggtgcg gggaaccagg gccaaacgtc cccgggtgaa tgtttcccgc 540

catcaaatta aaacacacag 560

<210> 153
<211> 577
<212> DNA
<213> Homo sapien

<400> 153
tgatgatata tggggcatgg tcctctagat gctgctcgag cggcgcagtg tgatggatgc 60

gtggtcgcgg cgaggtacca cctgttcatt ggggaactgt gggaaacgga gccaacggac 120

ctaagtgccc tttgacagtg agtttcatac catttcagta gtgtatttct ttcttaatct 180

gaataaacca gtatgatact ctcagacaca gaagaataaa gggagcgagt cattaacgtt 240

ttctttttaa acctttatga tgacttcctt atgaattact gaacgaacac tggaatggga 300

ctcaggtatc ctgaggacat ctctcaactc tggccttagt tccccctctg taaaattagg 360

gtgccaacta aatgatctac aaggtccctt ccagcgccgc cattctgtaa ttacatcatg 420

tgtaactgta ttaaacatac acaagtgact gccaggcatg ggaatgtaac ttccgagtaa 480

atgctttggt ttgttcagaa tacactatga acttctttcc aaagacgggt tgtggtaaat 540

agtggatatt ttgattataa gaaatagagt ttccttg 577

<210> 154
<211> 1138
<212> DNA
<213> Homo sapien

<400> 154
aagaaattcg gcacgaggaa agtgctggga ttacaagcat gagcccagcg cctggctgta 60

tctttcattt tacccaagtc actttaccca agtaagtaat taggggaaag cctgagtctt 120

gtaccacctg ttcatttggg gaactgtggg aaacggagcc aacggaccta agtgcccttt 180

gacagtgagt ttcataccat ttcagtagtg tatttctttc ttaatctgaa taaaccagaa 240

tgatactctc agcacagaag aataaaggga gcgagtcatt aacgttttct ttttaaacct 300

ttatgatgac ttccttatga attactgaac gaacactgga atgggactca ggtatcctga 360

ggacatctct caactctggc cttagttccc cctctgtaaa attagggtgc caactaaatg 420

atctacaagg tcccttccag cgccgccatt ctgtaattac atcatgtgta actgtattaa 480

acatacacaa gtgactgcca ggcatgggaa tgtaacttcc gagtaaatgc tttggtttgt 540

tcagaataca ctatgaactt ctttccaaag acgggttgtg gtaaatagtg gatattttga 600

ttataagaaa tagagtttcc ttgaagcttt agctggagat acagcaatag tgtggtgttc 660

ctacaaatat cacagtgtat tcaaacatat ttttctatca aaaatcattt ttgtaaaagc 720

tgtgtgtttt tatccaactt gtgataataa atgttcttta ttttagaata aaaaaaaaaa 780

aaaaaaaaaa aaagaaaaaa aaaggaaata aaaaaaaaaa acaggagaca aagacaacgg 840

cggcacgcaa caaccacatc gcggaaggcg acaagcgaac aacccagccc gagctcgtga 900

aggcgagcca acatgaagga gcgcactatc caagacaggt agctgacata acagaagaga 960

acaaaaacaa gagacaagta gaacaaaaac aaagagaaga caggacacac gagaaaagca 1020

ggtgtaatca gacgaacgac gcgacaaaca gagagacgtg caagcataaa atagcaacaa 1080

ccaagagaca gcgacggaca cacgaagcaa gacgagcgac gccgagcaca gcagggat 1138

<210> 155

<211> 800
<212> DNA
<213> Homo sapien

<400> 155
cgtggtcgcc ggccgaggtt gccataggcc ccagaccaaa ctagaccacc agcatgttca 60

tgtccagacc tcggcagtgg cgtgcactgc ttgtgcacct cagttcctcc agtgttggtt 120

tgtttgtttt ttaattcagc atcctgctgg ttttactttc caagcaagat ctgttgcgac 180

tcccaaatgc gttttaatga gctcatcctt atttgccttt cttcttacgt attttgttgt 240

atttaaagat tgtgcaggag atattctaga aggcattaat ggtttgcatt caaaacgatg 300

tggtttgtcc aagttatttt ctgtctttat tactgagacg gattaatctc cttatttttt 360

tcttgatgat ttgaagttgt aacagttgtc cagctattgc ttaataaaat tttgcagatc 420

aaaaaaaaaa aaaaaaaaaa aaaaaaggtt gggggtaacc agggccaaga ggggtccctc 480

ggggtcgaca attgggtcac ccgggtccat caatttcccc acaaacataa tacaggacat 540

aggcacacac agcaaacgca cacagcacca agacagacaa ctacggcgag ctaaggacgc 600

agagaagacg cggcaacgcg gaacgccccg agcaaggccg aggcaacaca ggagaggggc 660

agcgcacgac ggccggagca cgagcaggaa agcaacgaag agagacaacg gacacacgcg 720

agggcgaaga gaagagagca ggaacgacag gacaagcaca caaacgagcg gcaacagcag 780

acccagacga aacagcgcga 800

<210> 156
<211> 4632
<212> DNA
<213> Homo sapien

<400> 156
atgtatgcag cagtggaaca tgggcctgtg ctttgcagtg actccaacat cctgtgcctg 60

tcctggaagg ggcgtgtccc caagagtgag aaggagaagc ctgtgtgcag gagacgctac 120

tatgaggaag gctggctggc cacgggcaac gggcgaggag tggttggggt gactttcacc 180

tctagtcact gtcgcaggga caggagtact ccacagagga taaatttcaa cctccggggc 240

cacaatagcg aggttgtgct ggtgaggtgg aatgagccct accagaaact ggccacgtgc 300

gatgcggacg gaggcatatt cgtgtggatt cagtacgagg gcaggtggtc tgtggagctg 360

gtcaacgacc gcggggcgca ggtgagtgat ttcacgtgga gccatgatgg aactcaagca 420

cttatttcct atcgagatgg gtttgtcctg gttgggtctg tcagtggaca aagacactgg 480

tcatccgaaa tcaacttgga aagtcaaatt acgtgtggca tatggactcc tgacgaccaa 540

caggtgctgt ttggcacggc cgatgggcag gtgattgtca tggattgcca cggcagaatg 600

ctggcccacg tcctcttgca cgagtcagac ggtgtcctcg gcatgtcctg gaactacccg 660
atcttcctgg tggaggacag cagcgagagc gacacggact cagatgacta cgcccctccc 720
caagatggtc cggcagcata tcccatccca gtgcagaaca tcaagcctct gctcaccgtc 780
agcttcacct cgggagacat cagcttaatg aacaactacg atgacttgtc tcccacggtc 840
atccgctcag ggctgaaaga ggtggtagcc cagtggtgca cacaggggga cttgctggca 900
gtcgctggga tggaacggca gacccagctt ggtgagcttc ccaatggtcc ccttctgaag 960
agtgccatgg tcaagttcta caatgttcgt ggggagcaca tcttcacact ggacactctc 1020
gtgcagcgcc ccatcatctc catctgctgg ggtcaccggg attcgaggct gttgatggca 1080
tcaggaccag ccctgtacgt ggtgcgtgtg gagcaccggg tgtccagcct gcagctgctg 1140
tgccagcagg ccatcgccag caccttgcgt gaggacaagg acgtcagcaa gctgactctg 1200
ccccccgcc tctgctccta cctctccact gccttcatcc ccaccatcaa gcccccaatt 1260
ccagatccga acaacatgag agactttgtc agctacccat cagccggcaa cgagcggctg 1320
cactgcacca tgaagcgcac agaggacgac ccggaggtgg gcggcccgtg ctacacgctc 1380
tacctggagt acctgggcgg gcttgtgccc atcctcaaag ggcggcgcat cagcaagctg 1440
cggccagagt tcgtcatcat ggacccgcgg acagatagca aaccagatga aatctatggg 1500
aacagcttga tttctactgt gatcgacagc tgcaactgct cagactccag tgacattgag 1560
ctgagtgatg actgggctgc caagaaatct cccaaaatct ccagagctag caaatcaccc 1620
aaactcccaa ggatcagcat tgaggcccgc aagtcaccca agctgccccg ggctgctcag 1680
gagctctccc ggtccccacg gttgcccctg cgcaagccct ctgtgggctc gcccagcctg 1740
actcggagag agtttccttt tgaagacatc actcagcaca actatcttgc tcaggtcacg 1800
tctaatatct ggggaaccaa atttaagatt gtgggcttgg ctgctttcct gccaaccaac 1860
ctcggtgcag taatctataa aaccagcctc ctgcatctcc agccgcggca gatgaccatt 1920
tatctcccag aagttcggaa aatttccatg gactatatta atttacctgt cttcaaccca 1980
aatgttttca gtgaagatga agatgattta ccagtgacag gagcatctgg tgtccctgag 2040
aacagcccac cttgtaccgt gaacatccct attgcaccga tccacagctc ggctcaggct 2100
atgtccccca cgcagagcat agggctggtg cagtccctac tggccaatca gaatgtgcag 2160
ctagatgtcc tgaccaacca gacgacagct gtagggacag cagaacatgc aggtgacagg 2220
tgccacccag taacccaggt ctccaaccgg tactccaatc ctggacaggt gattttcgga 2280
agcgtggaaa tgggccgcat cattcagaac ccccctccac tgtccctgcc tcccccgccg 2340

caggggccca tgcagctgtc cacggtgggc catggagacc gagaccacga acacctgcag 2400
aagtcagcca aggccctgcg gccaacaccg cagctggcag ctgaggggga cgcagtggtc 2460
tttagtgccc cccaggaggt ccaggtgacg aagataaacc ctccaccccc gtacccagga 2520
accatccccg ctgcccccac cacagcagca cccccgcccc ctctgccgcc cccacagccc 2580
ccagtggatg tgtgcttgaa gaagggcgac ttctccctct accccacgtc agtgcactac 2640
cagaccccc tgggctatga gaggatcacc accttcgaca gcagtggcaa cgtggaggag 2700
gtgtgccggc cccgcacccg gatgctgtgc tcccagaaca cctacaccct ccccggcccg 2760
ggtagctctg ccaccttgag gctcacggcc actgagaaga aggtccctca gccctgcagc 2820
agtgccaccc tgaaccgcct gaccgtccct cgctactcca tccccaccgg ggacccaccc 2880
ccgtatcctg aaattgccag ccagctggcc caggggcggg gggctgccca gaggtccgac 2940
aatagcctca tccacgctac cctgcggagg aacaaccgtg aggctacgct caagatggcc 3000
cagctggccg acagcccgcg ggcccccctg cagcccctgg ccaagtccaa gggcgggccc 3060
gggggggtgg tgacacagct cccagcgcgg cccccacctg ccctgtacac ctgcagtcag 3120
tgcagtggca cagggcccag ctcacagccc ggagcctccc tggcccatac cgccagcgcc 3180
tccccgttgg cctcccagtc ctcctacagc ctcctgagcc cacccgacag cgcccgcgac 3240
cgcaccgact acgtcaactc ggccttcacg gaggacgagg ccctgtccca gcactgtcag 3300
cttgagaagc ccttgaggca ccctcccctg cctgaagctg ctgtcaccct gaaacggcca 3360
ccccccttacc agtgggaccc catgctgggt gaggacgttt gggttcctca agaaaggaca 3420
gcacagactt cagggcccaa cccccttaaaa ctgtcctctc tgatgctgag tcagggccag 3480
cacctggacg tgtcccgact gcccttcatc tcccccaagt ctcctgccag ccccactgcc 3540
actttccaaa caggctatgg gatgggagtg ccatatccag gaagctataa caaccccccct 3600
ttgcctggag tgcaggctcc ctgctctccc aaagatgccc tgtccccaac gcagtttgca 3660
caacaggagc ctgctgtggt ccttcagccg ctgtacccac ccagcctctc ctattgcacc 3720
ctgcccccca tgtacccagg aagcagcacg tgctctagtt tacagctgcc acctgtcgcc 3780
ttgcatccat ggagttccta cagcgcctgc ccgcccatgc agaaccccca gggcactctc 3840
cccccaaagc cacacttggt ggtggagaag ccccttgtgt ccccaccacc tgccgacctc 3900
caaagccact tgggcacaga ggtgatggta gagactgcag acaacttcca ggaagtcctc 3960
tccctgaccg aaagcccagt cccccagcgg acagaaaaat ttggaaagaa gaaccggaag 4020
cgcctggaca gccgagcaga agaaggcagc gttcaggcca tcactgaggg caaagtgaag 4080
aaggaggcta ggactttgag tgactttaat tccctaatct ccagcccaca cctggggaga 4140

gagaagaaga aagtgaagag tcagaaagac caactgaagt caaagaagtt gaataagaca 4200

aacgagttcc aggacagctc cgagagcgag cctgagctgt tcatcagcgg ggatgagctc 4260

atgaaccaga gccagggcag cagaaagggc tggaaaagca agcgctcccc acgggccgcc 4320

ggcgagctgg aggaggccaa gtgccggcgg gccagtgaga aggaggacgg gcggctgggc 4380

agccaaggct tcgtgtacgt gatggccaac aagcagccgc tgtggaacga ggccacccag 4440

gtctaccagc tggacttcgg ggggcgggtg acccaggagt ccgccaagaa cttccagatt 4500

gagttagagg ggcggcaggt gatgcagttt ggacggattg atggcagtgc gtacattcta 4560

gacttccagt atccgttctc agccgtgcag gcctttgcag ttgccctggc caacgtgact 4620

cagcgcctca aa 4632

<210> 157
<211> 998
<212> DNA
<213> Homo sapien

<400> 157
tgctgctcga gcgcgcgcag tgtgatggat ccgcccgggc aggtaccttt tcctctcaca 60

ttggcagaat agcacgcact agatgcctga ccttgagctc tagtctcccc gtttaaatct 120

taccttgggc agtaacgaca attattcctc attcaagtaa tttcaatgct gaaactgaac 180

tctattacta atgccttcca atcagagttc ctgatgggga tgcctgtggg atggcccact 240

aacctggggg acctaggcta gcatggggtg agttgggtaa ggaagatgat gcgttagttc 300

ctgatagatg ctacgagatg tagtttggca tttcagttgt tgtccagtta tgattttcac 360

tgggggttct gcagtcacag caagctgtgt atgaactagc tgtactagtg gatgacacac 420

tataactaat caaactagac taaagacaca ctgaaaatct gcgttataac taacaagata 480

tcactcatct gacacataac caccattaca ccttatggta cgtcaggatt cataaatagt 540

actgctctga atgacttatg ggaaatggtg ccactcaaaa gcaacttcct aacttgagga 600

ataactcctt tgtagtttac tttctggtac tggttggtgc cttgtatcgg gatacagcta 660

tattcttagc tcaaatgtct cttttggaga gcacagtagt tatcctactg gtgagactga 720

gaacctgagc tcatgagagg ccattccttc ctgggtgtcg gaccagggct ctgtgtcagg 780

aaaaaccttc tgggtgacct ttgtagactc gtttcaggtt gattccctct tatcttgcga 840

gagtagaatt cgcagtcggt ggcctttccc ttcacccgta actcggcccc tctgggcagg 900

ggcggggtgg cggctcttaa cgctggctcc gggtttgggg ggcccggggc ccgcaacgcg 960

ggttttgggc gggtcgcgcc cctccctcca acgggccg 998

<210> 158
<211> 766
<212> DNA
<213> Homo sapien

<400> 158
gggatgatcg ctcactatag ggcgctggtc actagatgca tgccgagcgg cgccaggtga 60

tggatcgagc ggccgcccgg gcaggtacat gttcatgaat ttgtgctgaa taattacttg 120

agtgtgaaat tgttatgtta tgcgatatat agtagtcaaa tatagaagat aatgcaaaac 180

aatttaaagt gattgtagca gttcgctgta ttctacagca gcaggattgt aggcagatta 240

ctgtagttct cacagcgagc agcatgtgag attggccagt ccgctcaaat tcgtgccaat 300

acttggtata tgctatcttg tcaatttcta gacattctgg agagtgtgta gtacttgttc 360

atcttggaca aattacactt aatagttatg tatccatttc tctaattttg ataacatttt 420

acataagttt atcgttatga gatatgttct ttattttgaa gtgcttattg tccattttac 480

attgggtcat ctgttattga attgtaaaca ttccttgaat atttaaatat gagtgcttgg 540

tcagttttgg tcacaaatat cctcgttttt tcacttttg cccttttatt attctgaaaa 600

tgccaagtga ttaaaattaa ttttactatt gttcaataaa caaaacaaaa aaaaaaaaaa 660

aaaaacacaa aaaaacaaaa gcgcgggggg taaccggggg cccaaggggg tcccccgggg 720

acattggtct ccccggtcac aattcccccc aatcgcacaa cagggc 766

<210> 159
<211> 1400
<212> DNA
<213> Homo sapien

<400> 159
ctatgattag cttattaggc tttgtggttt atatgcatca gaaagagtaa gacttaattt 60

tgtgtggaac aaataccctg gtgtagcatg tttcattaga atttgtttat agagatattg 120

ccatagaaaa gttatttttt attagtaaag aatgctttgt atttcctttg tggcttctaa 180

gtaccctttt ttggttatta tacctttatc cataagtatc tttaaatatt acaaaaatta 240

catattcttt taaatatttt aaagatttat tatattcatt taggttttaa tccactttta 300

attttttaga tgaaaagtaa gagaaaagta tataaatcat gagcacaaat tgaactaacc 360

aaggtaacaa tcaatctgct caagaaattg agcatcacca ccacctcctc ctgcactgtc 420

caaatcagca ccccagtact ccaaagcaaa tgttactcac tacactgact tctaacacaa 480

tagacttgtt ttgtctgttt tcaactatac aaaaatgaat catagagtat gtgttgtttt 540

gtatctggct cctttcacta aaattttggt ttataaaatt catccatgtg gttgaacaca 600

gttgtagatt gttcatttta attgttttac agtatttatt gtgtgactaa aacactactt 660

atttattcta taattgacag actttgggtt gcttttgctt tgggagtata aacattttta 720

tatctatgct ttaggtacat gttcatgaat ttgtgctgaa taattacttg agtgtgaaat 780

tgttatgtta tgcgatatat agtagtcaaa tatagaagat aatgcaaaac aatttaaagt 840

gattgtagca gtttgctgta ttctacagca gcagattgta gcagattact gtattctaca 900

gcagcagcat gtgagattgc cagttgctca aattcgtgcc aatacttggt atttttatc 960

ttttaatttt agacattctg gagagtgtgt agtaattttt catcttggaa aattacatta 1020

aattagtatc catttctcta attttgataa cattttcata agtttattgt tattagatat 1080

tttctttatt ttgaagtgct tattgtccat tttacattgg gtcatctgtt attgaattgt 1140

aaacattcct tgaatattta aatatgagtg cttggtcagt ttttgtcaca aatatcctct 1200

tttttcactt tttgcccttt tattattctg aaaatgccaa ttgattaaaa ttaattttac 1260

tattgttcaa taaacaaaac aaaaaaaaaa aaaaaaaaac acaaaaaaac aaaagcgcgg 1320

ggggtaaccg ggggcccaag ggggtccccg ggggacattg gtctccccgg tcacaattcc 1380

ccccaatcgc acaacagggc 1400

<210> 160
<211> 556
<212> DNA
<213> Homo sapien

<400> 160
acctattcac cattccaacg tgaagaagct ctgcatgtag gaaagaataa ttaacacact 60

tatagtctac tgcccatgta aggatcagct ccggctaaga ggccaaagat gggtgacatc 120

gtcatgctct gccttttatt ttttctttct tacccactta gcttcctaat tggaggaagg 180

aggcgtggta aaggtatatg aagactatgg tttaattaga ccagaaaaca ctgtcataat 240

ctctgggcgt cagtcagaat gtccagtttt gtctttgggc caagataagg gcagtgggat 300

ttatgatgtg ttgtttatag tctgaaacta ctctggtgat caccagggtc agtttcttta 360

atcgatggtt tccaagctgg cctaagtaca tttaagtaga gactgggctg ataaacatga 420

ccagacgaga cataaagacc ctgttgggaa tgacattgaa ctctcaaagt caagatttct 480

tacacaaatc tatcagctgg agaataatga gaggcagctg tggtatatgt gtgcaaataa 540

ggacattatg aagctt 556

<210> 161

<211> 1327
<212> DNA
<213> Homo sapien

<400> 161
ggaagacctg attgggaata gtcgaaagcc ttgatatgtg caaagaaaga accatttgat 60

caacccagtt cttaatacag gatactaact taaaatatag actcaagtta tacgataatt 120

caaacattta ttgtatttat actattctat atgtactttt ccaggaacca ggaatacaaa 180

actgacatgt tctctgtaca gaggctcaga ctagtagaga acagttaggt acgccgttaa 240

ttataaacta atatgtatca tcaattatgg gtttttatgg gggtttggca ggtggaaggg 300

accagggaga gatgatgagt gatgatggtt atgtagtctt taggaggatg caattataac 360

attgctcttc ctttcacgca ccacatgatt tagcaagtac ttcatattgg ctccaccatt 420

aacatggtca atggcttctg gatactcaca gttcaggcac agtttctcct gaagattttt 480

tacctctccc atctttaaga aattgtctgg atgtccatga aagatgctga cacttgtatt 540

aattcattaa aaaacaccac cccctccctg aaataaacta aaaagtaatg aattcataga 600

aaaaaatttc accaagattg aaactagaga atatacctag acttgcactt tgagctttga 660

gaaatgtgta cctattcacc attccaacgt gaagaagctc tgcagtagga aaaataatta 720

acacacttat agtctactgc ccatgtaagg atcagctccg gctaagaggc caaagatggg 780

tgacatcgtt atgctctgcc tttatttttt ctttcttacc cacttagctt cctaattgga 840

ggaaggaggc gtggtaaagg tatatgaaga ctatggttta attagaccag aaaacactgt 900

cataatctct ggggtcatca gaatgtccag ttttgtcttt gggccaagat aagggcagtg 960

ggatttatga tgtgttgttt atagtctgaa actactctgg tgatcaccag ggtcagtttc 1020

tttaatgatg gtttccaact ggcctaatac attaagtaag actggctgat aacatgacca 1080

gacagacata aagaccctgt tgggaatgac attgaactct caaagtcaag atttcttaca 1140

caaatctatc agctggagaa aatgaaggca gtgtggtata tgtgtgcaaa taaggacatt 1200

atgaagctta aatatggaat gtctcttgga cccccgatgt catctgtatt ctctttttct 1260

tcttgtacta tatcctttgc ctgtaaataa aaggtttatt tgaaaaaaaa aaaaaaaaaa 1320

gatcggc 1327

<210> 162
<211> 318
<212> DNA
<213> Homo sapien

<400> 162
ggttctccta aatgtcttaa cccatgttta tcttgttctg ctattccatg agcaaagaga 60

ataaagcaca aagctgtgag agtattaaat atggacacta gatttacatt tccaacaaga 120

aattcatctc cctccaaagt cccagaccag ggctagaatg tggttcattt ttaacaatca 180

aagtggcaag atctgtttgg tgatcactgt aaaacaggaa acacagtaat gccttcatgt 240

tgaggtgcta aaaggtcaag cttgggtaac aatgtccata gctgttctgg tgaatgtttc 300

gtcaatcaaa tagtgaaa 318

<210> 163
<211> 1042
<212> DNA
<213> Homo sapien

<400> 163
acagtctgtt tcctccttca cccccagaac aaaaatcgaa cttctggttg gacagtgtca 60

gatgtcactg aggtgacccc agcctgtttg cagttccaag tcttccgtgt aggcgtcact 120

gctactggaa ctttgtagat gaggagcctg tatgatgatg tcctgaacat ttctatcctt 180

tcctcacaca gagggaagct actgggaata tcagagacaa gctattatta aacaagtgtc 240

tctagtccaa gacatctcct gtggcaggga aatgaggggg caggctgtat cagtgatatt 300

tttataaact ctggttttag aaaaaattct tcagatggac gcattatttt aagactttaa 360

cattttccaa aaccaactga atcttatccc ctccatttat ccccctccag acacttctaa 420

tcaaggtcac catctccaac ttcccccata gacagtaaaa atatggctgg agaattctac 480

tgtaatagaa aaccaaggag atatggtaat ttgacagtgt gtttcctttc catccactag 540

acaagaatac cccctcccat tctttcctcc cctcagtcac cagaatgaag tgggctggaa 600

aacagttggt ctggttcctt tatagagact gattcccaca ttggatactg cctggaggcc 660

ttggggatga atgagaagtt ctgctggttt ggatcagtag cagaagcagg taacacatca 720

gggaaccggt cagcctaaga taggagggga cagaaaatga tgaaagagtt tctgatacat 780

ttatcagcta aattgctatg gtcaccccca tgtctcctgt aatgtccaac actaaggaat 840

taaactaagt aaactacaac ctttgtgtct tgctctgacc ttggaccaat ggaatatact 900

tcttatttca tattcagtgg ataagcaaat ctgcttcatc cctgccttaa ctcactcaag 960

gtctctgtga tgcactccag agttttcctc cttccctgca tagtcttctc ctccctagct 1020

gcctttcaaa ttggtgaaaa tg 1042

<210> 164
<211> 1120
<212> DNA
<213> Homo sapien

<400> 164
gccgcctttt tttttttttt tttttagaca agaaattatt ttagtccttt agtacagtct 60

gtttcctcct tcacccccag aacaaaaatc gaacttctgg ttggacagcg tcagatgtca 120

ctgaggtgac cccagcctgt ttgcagttcc aagtcttccg tgtaggcgtc actgctactg 180

gaactttgta gatgaggagc ctgtatgatg atgtcctgaa catttctatc ctttcctcac 240

acagagggaa gctactggga atatcagaga caagctatta ttaaacaagt gtctctagtc 300

caagacatct cctgtggcag ggaaatgagg gggcaggctg tatcagtgat atttttataa 360

actctggttt tagaaaaaat tcttcagatg gacgcattat tttaagactt taacattttc 420

caaaaccaac tgaatcttat cccctccatt tatccccctc cagacacttc taatcaaggt 480

caccatctcc aacttccccc atagacaata aaaatatggc tggagaattc tactgtaata 540

gaaaaccaag gagatatagt aatttgacag tgtgtttcct ttccatccac tagacaagaa 600

taccccctcc cattctttcc tcccctcagt caccagaatg aaggggctgg aaaacgttgg 660

tctggttcct tttagagctg attccccatt ggatactgcc tggaggcctt ggggatgaat 720

gagaagttct gcagtttgga tcagtagcag aagcaggtaa cacatcaggg aaccggtcag 780

cctaagatag gaggggacag aaaatgatga aagagtttct gatacattta tcagctaaat 840

tgctatggtc acccccatgt ctcctgtaat gtccaaacct aaggaattaa ctaagtaaac 900

taaaaccttt gtgttcttgc tctgaccttg gacaatggaa ttcttcttat tttcattcag 960

tggatagcaa atctgcttct tccctgcctt aactcactca aggtctctgt gatgcactcc 1020

agagttttcc tccttccctg catagtcttc tcctccctag ctgcctttca aattggtgaa 1080

aatgaagctt caggattatg aaaactagta cttaatgaag 1120

<210> 165
<211> 810
<212> DNA
<213> Homo sapien

<400> 165
agatcatgct cgagcggcgc agtgtgatgg attggtcgcg gccgaggtac ttttttatgg 60

cttacatctg tgcctggtcg gccatcaagt ctgggtgcca ctgtttgaga tttggggctg 120

tttcctgcaa ctgatctctg ctacagataa ggcttccctc ctggaggcca aagccctggt 180

taacgttaag agctctatga tgatgcaaac ttcagaggcg atcacctaac ataacaaaaa 240

cctccccaga accagaacct gtttttttcac caaaaccctt ccgctgcttg aataagaatg 300

tcttttcctt tcctaccaac tttgatgcca ctggccactg tgacataact tttacttagc 360

gggg taaatc atagatggat tacttgaact gccaacacaa gactgctgga cgagggacag 420

agctggatat gttagacaaa gatatacgaa cgacttggcg taatcactgg tcaatagctg 480

acaccatgat gtgaaaagta gtaatcacgg ctcacaagta ccaacacaag atacagaaga 540

caggagaaga ggaacaggaa aagaagaaac aacagagcac aaagagagaa caagcacaca 600

acagacgaag gccacaagag cgaaggagga ccggacgcag caccagcaac agaggaacgg 660

cacgcacaga agaacacaga caagaaaacg agaagaaacc acacgcacaa ctagccagaa 720

tcagagacag aaaacgcgaa gacaggaggc agaagcagaa acacaagaaa accgaacacc 780

aaaacaggca gcacaaacac gaagagaaag 810

<210> 166
<211> 601
<212> DNA
<213> Homo sapien

<400> 166
gaagtataac tatatgggcg aatgggtcct tagatgcagg ctcgagcggc gcagtgtgat 60

ggatccgccc gggcaggtac tcaggtgtta tatgattttc tgagctgaat aagtgcgagg 120

agcagattat taagatctgc cattctgaaa cgctggtctt tttctccttc ctatagtgca 180

ccataaaatt ctgttgatca gattatatta catacatttg ggggagtgga gggacatgag 240

ttaagtagcc cttcatgtat ttataatctc ttttctactg aatcaaatga cttagccatg 300

accctgaatg gacctgtttt acttcaagtg agatgtctgc cttttatgaa ttgtatatgt 360

gaatagagtt cgggggttgc caaaaatgca tacatgtatg taagtaaaat tttttatgaa 420

gtagtctgtc aaatgtatca taaagtttat ttttctttta tacgtaaatc attaaaaata 480

atcacatatt tttgaaaaaa aaaaaaaaaa aaaaaaaggt gggggtatc tcggggccaa 540

aaggggtccc gggggggaat tggttttccg gttcaaattt ccacaaattt gggagaaaat 600

a 601

<210> 167
<211> 1035
<212> DNA
<213> Homo sapien

<400> 167
tggtcgcggc gaggtactgt aaatgtgatg gaaaacattg atgagaattt attggcagtt 60

cagattgtgt tttcccaact taggctcttt attaattggt taaggttttc tccaaaaagg 120

gcatttcaac aatgggaatt attttaaatt ggttaaacca gtgggcacag attacttatc 180

ttccttctct gctttgtgac tcaccagcag taacacacac aatccacatc ttgtgcacct 240

caaatgaaca gacttggttt ccttgctttc ttgacatttc catgactgtt tcacatacaa 300

actattgggt gaggtttttc agctgttacc gacccacgtc ctgctgtctc tgtgtggtcc 360

tacaaaaact gtccattccc accccttttgc tttgccattt gcaagagtct ggaattgtca 420

ggtctcagct tcgaaaagtc ctggttccac tgacaggaca cattctttag tgggaattaa 480

gacctacaaa gtctagtttg tatgtaggta tgaagggaat tttttaaata aagtggaaaa 540

gctgtgaaca gcattagaac tctgtctatt tcttaatttt aaaatatgct gatatgcctt 600

aaactgtagt tgtagatcct tgtcattttg ctgtttgaaa ataaccaatg tgttttctaa 660

aactgtcgtg taatctactt tcattgttaa tgcagaattg tcatatatgt aagccgcatg 720

ttagacattt gtctttttta aactaaagta attgtattga tgtgaagcat atcatttttt 780

caaatatgaa agtgatcact tagcaacatg cttggtaatt tggcatctgt taaggtagga 840

gagtggtgaa cagataatct atgcatatat cactagtgcc aagacataaa gcgggggaaa 900

atatattttt acccaaacat taaaaaaaac aaaaaaaaaa aaaaaaaaaa aaaaaaaggc 960

tgggggtaac cggggccaaa ggggtcccgg ggtgaattgg ttttccgctc aaattccccc 1020

atttttgggc aaacc 1035

<210> 168
<211> 1666
<212> DNA
<213> Homo sapien

<400> 168
ctgggtgatg aagtgagact ctccaaaaaa aaaaagaaat tattaatccc tgcctgtgct 60

ctacatagcc tcatgggcat cattggatag ctcagagggc ccttgattct ggcaaggcaa 120

ataaagccag aatgagaaat taccatcttc tactagagaa aaccaagaga aaaatttttta 180

tgctaggatg cctttatgac cacttaattt tttaatctta gtttaatggt ctctccctgg 240

tgctaactgc tgacagtggc cacctctttt ttggggattg aggggcctac ataactagct 300

ggccttaccc catatctttt gttcaaacat aataccatct ttttgcttct tctgaacttt 360

agatctccat aacacatgta ctgtagaatg tgatggaaaa gcattgatga gaatttattg 420

gcagttcaga ttgtgttttc ccaacttagg ctctttatta attggttaag gttttctcca 480

aaaagggcat ttcaacaatg ggaattattt aatgtaacag tgggcacaga ttacttatct 540

tccttctctg ctttgtgact caccagcagt aacacacaca atccacatct tgtgcacctc 600

aaatgaacag acttggtttc cttgctttct tgacatttcc atgactgttt cacatacaaa 660

ctattgggtg aggtttttca gctgttaccg acccacgtcc tgctgtctct gtgtggtcct 720

acaaaaactg tccattccca cccctttgct ttgccatttg caagagtctg gaattgtcag 780

gtctcagctt cgaaaagtcc tggttccact gacaggacac attctttagt gggaattaag 840

acctacaaag tctagtttgt atgtaggtat gaagggaatt ttttaaataa attgaaaagc 900

tgtgaacagc attagaactt tgtctatttc ttaattttaa aatatgctga tatgccttaa 960

actgtagttg tagatccttg tcattttgct gtttgaaaat aaccaatgtg ttttctaaaa 1020

ctgtcgtgta atctactttc attgttaatg cagaattgtc atatatgtaa gctgcatgtt 1080

agacatttgt cttttttaaa ctaaagtaat tgtattgatg tgaagcatat catttttttca 1140

aatatgaaag tgatcactta gcaacatgct tggtaatttg gcatctgtta aggtaggaga 1200

gtggtgaaca gataatctat gcatatatca ctagtgccaa gacataaagc gggggaaaat 1260

atatttttac ccaaacatta aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa caactgtgtt 1320

cggcgcgctt gtggccccgg aagaagagtc ttctcgtaga accatcgtgg tttgggccca 1380

gcggggcccc aggaggtagg gtgccacacg ggccaaaagc gtgtcccagg agacacccgg 1440

gggcactaga acaacttagg gtgtgtgagg aatattttcg ctcaccccat gttacaaaaa 1500

caaccgcgca gagggggcaa acagcaacag ggtttctgtg aaacaacaac ccccaaatgg 1560

agggaagtcc tcgagaagga catacaggga aagcctaata caacagaggg aagatcccaa 1620

ggaaaagcac tatcatataa ataattatcg ccgccggctg tgcggg 1666

<210> 169
<211> 633
<212> DNA
<213> Homo sapien

<400> 169
aaaacaacac ggaatgtcta cgactaacta tagggcccct ggtgtatcta gatgcatgct 60

cgagccggcc gccatgatgt gactggatgt cgcggccgag gtacagagta tgtagtgggc 120

atctgttgaa tgaatgcttt tcccagtacg cacgtgtatt catacaatat taatataatt 180

agtcccctgg gcttacagat aaaaatgaaa cgcatcaacg tgcccagctg cagtgagacc 240

caggtgttct tcctccaccc ctagtggtcc cctgggcagg tctttttttt ttggtaacac 300

tcaccaggtc tgttctgtag tcaatcatgt gatggactgt gtcggtgaac tgtgcaggac 360

actgttctca tagtgttcat tagcgacaga gtaaacatgt ttgccatgca agggttattt 420

ggcatctgca tttaagtgat aatgttgaat caatgaaaag gtgttgatta agcagtagtt 480

gtagatatgc taagtttttc aaattactaa tatcaagtgg agatggtttt tactttataa 540

gggtattgct ttggtgatag cataaataat gggtttccct ttttggtaac tgtaacatta 600

attggctggc aactttggta ttcccataga ctg 633

<210> 170
<211> 563
<212> DNA
<213> Homo sapien

<400> 170
gggaaggaag acatataggg cggaatgggt cctagatgca tgtcgagcgg cgcagtgtga 60

tggatcggcg ccgggcaggt acaaaaaata ggataaatgc ttgtttttt atttagcaat 120

gtccaaaata atgaattgat ttcccgagta tcctctaaag gtaaccaggg attttttta 180

tttaattatc ttgaacccac atatttaaat atacgtagta tgctacaaac cattgcagtt 240

aagtaccttt attgatgctt gagttgccca ctttttcttt tttttttttt ggagacagag 300

cctcgctctg tcacccaggc tggagtgcag gggcgtcatc tttgactcac ttgcaacctt 360

ccttccttcc gtggggtgca ggcagattct cctgtgcctt acagcctccg agtttggctg 420

ggatttacag ggcattgttg caagtttccc acattttcag tgagaaattc ctcaattggc 480

ctccgtgagt ggtttggaaa ttgaccccag aattcttgga gtgggtgtat tagctatcta 540

tggctggtgt aacaaattga cct 563

<210> 171
<211> 682
<212> DNA
<213> Homo sapien

<400> 171
gaaaaggttg gcagcaggtg cacgtgttat cagcctgatc atctatcacc tgatggtttt 60

agcaatacct aaatccgtga tatcatcaga ggttgcaaaa tgatgagatt caggtttttt 120

ttttacataa ttattggtca gaattattct gcaaatagct tctctttaac agtattcggt 180

taccttgaaa tacaggttgt acaaaaaata ggataaatgc ttgttttttt atttagcaat 240

gtccaaaata atgaattgat ttcccagtat cctctaaagg taaccaggga ttttttttat 300

ttaattatct tgaacccaca tatttaaata tacgtagtat gctacaaacc attgcagtta 360

atacctttat tgatgcttga gttgcccact tttttctttt ttttttttg gagacagagc 420

ctcgctctgt cacccaggct ggagtgcagg ggcgtcatct ttgactcact tgcaaccttc 480

cttccttccg tggggtgcag gcagattctc ctgtgcctta cagcctccga gtttggctgg 540

gatttacagg gcattgttgc aagtttccca cattttcagt gagaaattcc tcaattggcc 600

tccgtgagtg gtttggaaat tgaccccaga attcttggag tgggtgtatt agctatctat 660

ggctggtgta acaaattgac ct 682

<210> 172
<211> 75
<212> PRT
<213> Homo sapien

<400> 172

Met Gly Pro Arg Ser Arg Leu Trp Pro Ser Ser Pro Leu Trp Leu Val
1 5 10 15

Gln Pro Leu Cys Thr Pro Gly Val Phe Thr Pro Gly Ala Asp Ser Ser
20 25 30

His Cys Ser Ser Phe Leu Arg Glu Ile Thr Val Val Ile Ala Ala Gly
35 40 45

Ala Asn Arg Leu Gly Leu Val Ser Cys Ala Phe Gly Gln Leu Leu Thr
50 55 60

Arg Ser Ser Leu Lys Gln Trp Gly Gly Pro His
65 70 75

<210> 173
<211> 38
<212> PRT
<213> Homo sapien

<400> 173

Met Phe Pro Arg Leu Asp Ser Thr Ser Trp Pro Gln Gly Ile Leu Trp
1 5 10 15

Ala Trp Thr Pro Lys Pro Leu Arg Leu Glu Val Cys Glu Pro Pro Ser
20 25 30

Leu Pro Ser Leu Trp Ser
35

<210> 174
<211> 52
<212> PRT
<213> Homo sapien

<400> 174

Met Thr Leu Phe Ile Arg Cys Cys Thr Asn Tyr Gly Asn Leu Cys Gln
1 5 10 15

```
Tyr Phe Asn Val Cys Trp Ile Ile Thr Asp Ile Phe Ile Ile Leu Met
            20                  25                  30

Ser Thr Asn Leu Phe Ile Leu Ile Ala Arg Val Ser Leu Gly Ser Lys
        35                  40                  45

His His Leu Gly
    50


<210>  175
<211>  37
<212>  PRT
<213>  Homo sapien

<400>  175

Met Ala Gly Ser Gly Lys Val Pro Ile Thr Thr Thr Tyr Lys Pro Pro
1               5                   10                  15

Thr Asn Ser Asn Ala Ile His Leu Pro Thr Pro Ile Ile Arg Lys Ala
            20                  25                  30

Gly Phe Thr Gly Ile
        35


<210>  176
<211>  88
<212>  PRT
<213>  Homo sapien

<400>  176

Met Gly Leu Thr Leu Lys Ser Leu Cys Asp Ser Lys Met Asn Cys Gln
1               5                   10                  15

Ser Asn Val Pro Leu Met Lys Asp Pro Ile Thr Leu Gln His Val Cys
            20                  25                  30

Ile Gln Arg Thr Tyr Leu Arg Leu Ser Phe Gly His Gly Gly Arg Leu
        35                  40                  45

Leu Leu Lys Thr Tyr Gln Ser Pro Leu Trp Arg Ser Ala Asp Arg Pro
    50                  55                  60

His Asp Leu Gly Asn Gly Leu Leu Val Ile Trp Asp Cys Leu Gly Leu
65                  70                  75                  80
```

Cys Asn Gly Thr Trp Gly Gln Asn
85

<210> 177
<211> 61
<212> PRT
<213> Homo sapien

<400> 177

Met Asp His Lys Ser Ala Asn His Ser Ser Ala Leu Leu Lys Met Leu
1 5 10 15

Leu Ala Gly Gly Met Ser Leu Pro Glu Val Pro Glu Gly Leu Thr Pro
20 25 30

Thr Pro Ser Ser Gln Thr His Leu Ser Lys Gly Lys Gly Arg Asn Leu
35 40 45

Glu Lys Ser Tyr Phe His Asn His Ser Leu Arg Glu Pro
50 55 60

<210> 178
<211> 198
<212> PRT
<213> Homo sapien

<400> 178

Met Thr Pro Ile His Leu Ile Cys Ser Pro Ser His Glu Leu Gln Asp
1 5 10 15

Thr Thr His Pro Gln Pro Gln Arg Glu Cys Gln Arg Phe Ser Thr His
20 25 30

Gly Ala Gln Thr Thr Gln Cys Ala Thr His His His Pro Tyr Ile Ser
35 40 45

Gly Ala Ala Thr Arg Thr Tyr Leu Arg His Val Ala Pro Asp Tyr Ser
50 55 60

Ala Pro Leu Met Ala Pro Pro Thr Asn Thr Arg Leu Ala Pro Ala Ser
65 70 75 80

Leu Gln Pro Thr His Leu Arg Pro Pro Leu Ala Arg His Pro Leu Thr
85 90 95

Ala Asp Cys Arg Thr His Gln Leu Thr Asp Thr Arg Pro Leu His Pro

100 105 110

Arg Pro Ile Thr Ser Arg Thr Pro Gln Pro Leu Pro Ser His Thr His
115 120 125

Gly Leu His His Thr Arg Pro Pro His Thr Ala Thr Gly Cys Pro Tyr
130 135 140

Leu Ser Thr Ser Arg Pro Leu Pro Pro Leu His Thr Arg Ser Ile His
145 150 155 160

Pro Asp Asn Pro His Cys Thr Thr Pro His His Ser Pro Ser Lys Pro
165 170 175

Ser Thr Thr Thr His Gln Gln Ser Pro Ala Pro Thr Pro Asn Lys Pro
180 185 190

His Pro Arg Arg Ala Ser
195

<210> 179
<211> 20
<212> PRT
<213> Homo sapien

<400> 179

Met Ile Gly Ile Thr Trp Cys Phe Glu Leu Ile His Pro Thr Leu Glu
1 5 10 15

Leu Thr Ala Thr
20

<210> 180
<211> 107
<212> PRT
<213> Homo sapien

<400> 180

Met Gly Ala Ser Gly Pro Glu Arg Glu Asp Arg Asn Ser Glu Asn Gly
1 5 10 15

Val Glu Lys Lys Asn Val Lys Glu Leu His Glu Glu His Met Ala Glu
20 25 30

Lys Lys Glu Leu Gln Glu Glu Asn Gln Arg Leu Gln Gly Leu Pro Val
35 40 45

Ser Gly Ser Glu Glu Gly Arg Leu Pro Val Pro Ser Ala Arg Ser Ser
    50   55   60

Thr Leu Arg Ala Ser Cys Arg Asn Glu Leu Gly Ser Leu Leu Pro Gly
65   70   75   80

Gly Glu Thr Ser Leu Gly Leu Lys Glu Gly His Arg Thr Lys Gly Ala
                85   90   95

Arg Gly Gly His Arg Glu Asp Pro Gln Glu Lys
            100   105

<210> 181
<211> 27
<212> PRT
<213> Homo sapien

<400> 181

Met Ser Thr His Ser Val His Ser Thr Gly Leu Pro Phe Tyr Lys Leu
1   5   10   15

Ser Leu Thr Ser Leu Ser Ser Met Thr Leu Val
            20   25

<210> 182
<211> 40
<212> PRT
<213> Homo sapien

<400> 182

Cys Phe Glu Lys Met Leu Asn Arg Leu Gly Ala Val Ala His Val Cys
1   5   10   15

Asn Pro Ser Thr Leu Gly Gly Arg Gly Gly Trp Ile Met Arg Ser Gly
            20   25   30

Val Arg Asp Gln Pro Gly Gln His
        35   40

<210> 183
<211> 26
<212> PRT
<213> Homo sapien

<400> 183

Met Arg Lys Gln Ala Phe Asp Leu Leu Glu Ser Thr Ala Gln Lys Ser
1 5 10 15

Leu Val Pro Ile Phe Glu Phe Pro Lys Gln
20 25

<210> 184
<211> 39
<212> PRT
<213> Homo sapien

<400> 184

Met Lys Glu Glu Gly Arg Leu Leu Thr Val Ala Glu Gly Arg Gln Gly
1 5 10 15

Pro Ser Cys Ser Ser His Ile Asn Ser Lys Lys Pro Ser Gln Gln Asn
20 25 30

Lys Ser Ile Phe Asn Ser Ser
35

<210> 185
<211> 76
<212> PRT
<213> Homo sapien

<400> 185

Met Val Glu Pro Ala Leu Ser Gly Cys Gln Gln Arg Lys Gly Gly Tyr
1 5 10 15

Ser Ser Glu Arg Gln Ser Gln Pro Thr Gln Gly Gly Gln Gly Val Arg
20 25 30

Pro Gln Thr Tyr Ser Pro Ala Asp Leu Thr Val Arg Pro Ser Cys Ser
35 40 45

Gly Thr Gly Asn Ala Gln Ala Glu Ile Ala Leu Leu His Thr Tyr Asn
50 55 60

Thr Thr Leu Glu Asn Asn Leu Glu Trp Phe Thr Leu
65 70 75

<210> 186
<211> 35
<212> PRT
<213> Homo sapien

<400> 186

Met Arg Gln Pro Cys Leu Ala Ile Pro Glu Ala Ser Ala Ser Leu Ile
1   5   10   15

Cys Arg Cys His Arg His Phe Thr Tyr Ser His Leu Met Ala Arg Phe
            20   25   30

Leu Leu Leu
        35

<210> 187
<211> 76
<212> PRT
<213> Homo sapien

<400> 187

Met Phe Phe Ala Leu Met Gly Ile Cys Pro Gly Thr Leu Pro Pro Gly
1   5   10   15

Pro Pro Leu Pro Arg Trp Pro Pro Pro Val Phe Cys Phe Phe Phe Phe
            20   25   30

Phe Phe Gly Phe Phe Phe Cys Cys Phe Thr Val Lys Leu Phe Ile Glu
        35   40   45

Gln Ile Glu Asp Asn Asp Ile Cys Phe Tyr Tyr Arg Ser Leu Pro Ser
    50   55   60

Ser Tyr Ile Ile Asp Thr Tyr Tyr Glu Thr Cys Ile
65   70   75

<210> 188
<211> 173
<212> PRT
<213> Homo sapien

<400> 188

Met Ile Gly Cys Ser Leu Leu Val Ala Cys Leu Cys Cys Leu Val Gln
1   5   10   15

Ser Phe Arg Ala Met Phe Ser Cys Phe Ser Gly Leu Ser Leu Cys Leu
            20   25   30

Met Leu Pro Leu Trp Cys Val Cys Pro Thr Val Cys Ala Phe Phe Cys
        35   40   45

```
Gly Tyr Leu Leu Phe Phe Ser Leu Arg His Ala Ala Cys Gly Cys Leu
    50                  55                  60

Leu Val Cys Leu Ser Cys Leu Ala Leu Pro Ser Gly Pro Ile Leu Ser
65                  70                  75                  80

Phe Ser Phe Cys Leu Arg Val Val Ser Ser Val Arg Val Ala Cys Ala
                85                  90                  95

Arg Ser Ala Ala Val Leu Leu Leu Arg Gly Val Pro Pro Pro Ser Leu
            100                 105                 110

Arg Thr Leu Ser Leu Ile Ala Ser Thr Ala Thr Arg Leu Ser Phe Val
        115                 120                 125

Phe Leu Phe Ser Leu Pro Arg Gly Leu Leu Cys Val Gly Gly Ser Gly
    130                 135                 140

Ser Val Leu Gly Ser Leu Val Arg Arg Ala Gln Ser Val Gly Leu Arg
145                 150                 155                 160

Asp Phe Val Ser Val Leu Gln Val Val Leu Thr Cys Leu
                165                 170


<210>  189
<211>  29
<212>  PRT
<213>  Homo sapien

<400>  189

Met Val Leu Tyr Ser Glu Gly His Gln His Gly Pro His Leu Leu Asn
1               5                   10                  15

Met Glu Asn Gln Asn Leu Asn Glu Leu Pro Leu Lys Gly
            20                  25


<210>  190
<211>  122
<212>  PRT
<213>  Homo sapien

<400>  190

Phe Phe Ala Asp Glu Val Ser Arg Leu Ser Pro Gly Leu Glu Cys Ser
1               5                   10                  15
```

Gly Val Ile Ser Ala His Cys Asn Phe His Leu Leu Gly Ser Ser Ser
20 25 30

Ser Pro Ala Ser Ala Ser Gln Val Ala Glu Ile Thr Gly Ala Cys His
35 40 45

Pro Thr Trp Leu Ile Phe Val Ile Leu Val Glu Thr Gly Phe His His
50 55 60

Val Gly Gln Ala Asp Ala Leu Leu Thr Ser Gly Asp Pro Pro Phe Ser
65 70 75 80

Ala Pro Lys Val Leu Gly Ile Thr Gly Val Ser His Arg Ala Arg Pro
85 90 95

Ala Asn Thr Phe Ala Leu Thr Thr Leu Gly Leu Leu Tyr Lys Ile Val
100 105 110

Met Ile Ala Met Glu Val Leu Pro Val Pro
115 120

<210> 191
<211> 11
<212> PRT
<213> Homo sapien

<400> 191

Met Trp Arg Ala Lys Gln Tyr Asp Leu Gln Thr
1 5 10

<210> 192
<211> 28
<212> PRT
<213> Homo sapien

<400> 192

Met Met Phe Ser Leu Ser Gln Lys Gly Ser Ala Ala Val Gln Ser Pro
1 5 10 15

Ser Thr Leu Ser Thr Pro Thr Phe Ser Ile Ser Tyr
20 25

<210> 193
<211> 48
<212> PRT
<213> Homo sapien

<400> 193

Met Asp Ser Gly Ala Arg Ala Gly Lys Pro Leu Leu Asp Pro Val Cys
1 5 10 15

Leu Pro Ala Trp Ser Leu Cys Leu Gln Pro Cys Leu Tyr Ser Ser Leu
20 25 30

Pro Pro His Gln Pro Pro Leu Ala Ser Pro Tyr Arg Leu Ser Lys Lys
35 40 45

<210> 194
<211> 1138
<212> PRT
<213> Homo sapien

<400> 194

Met Gly Asp Phe Ala Ala Pro Ala Ala Ala Ala Asn Gly Ser Ser Ile
1 5 10 15

Cys Ile Asn Ser Ser Leu Asn Ser Ser Leu Gly Gly Ala Gly Ile Gly
20 25 30

Val Asn Asn Thr Pro Asn Ser Thr Pro Ala Ala Pro Ser Ser Asn His
35 40 45

Pro Ala Ala Gly Gly Cys Gly Gly Ser Gly Gly Pro Gly Gly Gly Ser
50 55 60

Ala Ala Val Pro Lys His Ser Thr Val Val Glu Arg Leu Arg Gln Arg
65 70 75 80

Ile Glu Gly Cys Arg Arg His His Val Asn Cys Glu Asn Arg Tyr Gln
85 90 95

Gln Ala Gln Val Glu Gln Leu Glu Leu Glu Arg Arg Asp Thr Val Ser
100 105 110

Leu Tyr Gln Arg Thr Leu Glu Gln Arg Ala Lys Lys Ser Gly Ala Gly
115 120 125

Thr Gly Lys Gln Gln His Pro Ser Lys Pro Gln Gln Asp Ala Glu Ala
130 135 140

Ala Ser Ala Glu Gln Arg Asn His Thr Leu Ile Met Leu Gln Glu Thr
145 150 155 160

Val Lys Arg Lys Leu Glu Gly Ala Arg Ser Pro Leu Asn Gly Asp Gln
165 170 175

Gln Asn Gly Ala Cys Asp Gly Asn Phe Ser Pro Thr Ser Lys Arg Ile
180 185 190

Arg Lys Asp Ile Ser Ala Gly Met Glu Ala Ile Asn Asn Leu Pro Ser
195 200 205

Asn Met Pro Leu Pro Ser Ala Ser Pro Leu His Gln Leu Asp Leu Lys
210 215 220

Pro Ser Leu Pro Leu Gln Asn Ser Gly Thr His Thr Pro Gly Leu Leu
225 230 235 240

Glu Asp Leu Ser Lys Asn Gly Arg Leu Pro Glu Ile Lys Leu Pro Val
245 250 255

Asn Gly Cys Ser Asp Leu Glu Asp Ser Phe Thr Ile Leu Gln Ser Lys
260 265 270

Asp Leu Lys Gln Glu Pro Leu Asp Asp Pro Thr Cys Ile Asp Thr Ser
275 280 285

Glu Thr Ser Leu Ser Asn Gln Asn Lys Leu Phe Ser Asp Ile Asn Leu
290 295 300

Asn Asp Gln Glu Trp Gln Glu Leu Ile Asp Glu Leu Ala Asn Thr Val
305 310 315 320

Pro Glu Asp Asp Ile Gln Asp Leu Phe Asn Glu Asp Phe Glu Glu Lys
325 330 335

Lys Glu Pro Glu Phe Ser Gln Pro Ala Thr Glu Thr Pro Leu Ser Gln
340 345 350

Glu Ser Ala Ser Val Lys Ser Asp Pro Ser His Ser Pro Phe Ala His
355 360 365

Val Ser Met Gly Ser Pro Gln Ala Arg Pro Ser Ser Ser Gly Pro Pro
370 375 380

Phe Ser Thr Val Ser Thr Ala Thr Ser Leu Pro Ser Val Ala Ser Thr

```
385                 390                 395                 400

Pro Ala Ala Pro Asn Pro Ala Ser Ser Pro Ala Asn Cys Ala Val Gln
                405                 410                 415

Ser Pro Gln Thr Pro Asn Gln Ala His Thr Pro Gly Gln Ala Pro Pro
            420                 425                 430

Arg Pro Gly Asn Gly Tyr Leu Leu Asn Pro Ala Ala Val Thr Val Ala
        435                 440                 445

Gly Ser Ala Ser Gly Pro Val Ala Val Pro Ser Ser Asp Met Ser Pro
    450                 455                 460

Ala Glu Gln Leu Lys Gln Met Ala Ala Gln Gln Gln Gln Arg Ala Lys
465                 470                 475                 480

Leu Met Gln Gln Lys Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln
                485                 490                 495

Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln His Ser
            500                 505                 510

Asn Gln Thr Ser Asn Trp Ser Pro Leu Gly Pro Pro Ser Ser Pro Tyr
        515                 520                 525

Gly Ala Ala Phe Thr Ala Glu Lys Pro Asn Ser Pro Met Met Tyr Pro
    530                 535                 540

Gln Ala Phe Asn Asn Gln Asn Pro Ile Val Pro Pro Met Ala Asn Asn
545                 550                 555                 560

Leu Gln Lys Thr Thr Met Asn Asn Tyr Leu Pro Gln Asn His Met Asn
                565                 570                 575

Met Ile Asn Gln Gln Pro Asn Asn Leu Gly Thr Asn Ser Leu Asn Lys
            580                 585                 590

Gln His Asn Ile Leu Thr Tyr Gly Asn Thr Lys Pro Leu Thr His Phe
        595                 600                 605

Asn Ala Asp Leu Ser Gln Arg Met Thr Pro Pro Val Ala Asn Pro Asn
    610                 615                 620
```

```
Lys Asn Pro Leu Met Pro Tyr Ile Gln Gln Gln Gln Gln Gln Gln Gln
625                 630                 635                 640

Gln Gln Gln Gln Gln Gln Gln Gln Gln Gln Pro Pro Pro Pro Gln Leu
                645                 650                 655

Gln Ala Pro Arg Ala His Leu Ser Glu Asp Gln Lys Arg Leu Leu Leu
            660                 665                 670

Met Lys Gln Lys Gly Val Met Asn Gln Pro Met Ala Tyr Ala Ala Leu
        675                 680                 685

Pro Ser His Gly Gln Glu Gln His Pro Val Gly Leu Pro Arg Thr Thr
    690                 695                 700

Gly Pro Met Gln Ser Ser Val Pro Pro Gly Ser Gly Gly Met Val Ser
705                 710                 715                 720

Gly Ala Ser Pro Ala Gly Pro Gly Phe Leu Gly Ser Gln Pro Gln Ala
                725                 730                 735

Ala Ile Met Lys Gln Met Leu Ile Asp Gln Arg Ala Gln Leu Ile Glu
            740                 745                 750

Gln Gln Lys Gln Gln Phe Leu Arg Glu Gln Arg Gln Gln Gln Gln Gln
        755                 760                 765

Gln Gln Gln Gln Ile Leu Ala Glu Gln Gln Leu Gln Gln Ser His Leu
    770                 775                 780

Pro Arg Gln His Leu Gln Pro Gln Arg Asn Pro Tyr Pro Val Gln Gln
785                 790                 795                 800

Val Asn Gln Phe Gln Gly Ser Pro Gln Asp Ile Ala Ala Val Arg Ser
                805                 810                 815

Gln Ala Ala Leu Gln Ser Met Arg Thr Ser Arg Leu Met Ala Gln Asn
            820                 825                 830

Ala Gly Met Met Gly Ile Gly Pro Ser Gln Asn Pro Gly Thr Met Ala
        835                 840                 845

Thr Ala Ala Ala Gln Ser Glu Met Gly Leu Ala Pro Tyr Ser Thr Thr
    850                 855                 860
```

Pro Thr Ser Gln Pro Gly Met Tyr Asn Met Ser Thr Gly Met Thr Gln
865 870 875 880

Met Leu Gln His Pro Asn Gln Ser Gly Met Ser Ile Thr His Asn Gln
885 890 895

Ala Gln Gly Pro Arg Gln Pro Ala Ser Gly Gln Gly Val Gly Met Val
900 905 910

Ser Gly Phe Gly Gln Ser Met Leu Val Asn Ser Ala Ile Thr Gln Gln
915 920 925

His Pro Gln Met Lys Gly Pro Val Gly Gln Ala Leu Pro Arg Pro Gln
930 935 940

Ala Pro Pro Arg Leu Gln Ser Leu Met Gly Thr Val Gln Gln Gly Ala
945 950 955 960

Gln Ser Trp Gln Gln Arg Ser Leu Gln Gly Met Pro Gly Arg Thr Ser
965 970 975

Gly Glu Leu Gly Pro Phe Asn Asn Gly Ala Ser Tyr Pro Leu Gln Ala
980 985 990

Gly Gln Pro Arg Leu Thr Lys Gln His Phe Pro Gln Gly Leu Ser Gln
995 1000 1005

Ser Val Val Asp Ala Asn Thr Gly Thr Val Arg Thr Leu Asn Pro
1010 1015 1020

Ala Ala Met Gly Arg Gln Met Met Pro Ser Leu Pro Gly Gln Gln
1025 1030 1035

Gly Thr Ser Gln Ala Arg Pro Met Val Met Ser Gly Leu Ser Gln
1040 1045 1050

Gly Val Pro Gly Met Pro Ala Phe Ser Gln Pro Pro Ala Gln Gln
1055 1060 1065

Gln Ile Pro Ser Gly Ser Phe Ala Pro Ser Ser Gln Ser Gln Ala
1070 1075 1080

Tyr Glu Arg Asn Ala Pro Gln Asp Val Ser Tyr Asn Tyr Ser Gly
1085 1090 1095

Asp Gly Ala Gly Gly Ser Phe Pro Gly Leu Pro Asp Gly Ala Asp
1100 1105 1110

Leu Val Asp Ser Ile Ile Lys Gly Gly Pro Gly Asp Glu Trp Met
1115 1120 1125

Gln Glu Leu Asp Glu Leu Phe Gly Asn Pro
1130 1135

<210> 195
<211> 30
<212> PRT
<213> Homo sapien

<400> 195

Met Gln Leu Pro Leu Ser His Lys Arg Lys Lys Gln Tyr Ser Phe Tyr
1 5 10 15

Val Phe Asp Thr Asn Ile Lys His Asn Ser Val Leu Val His
20 25 30

<210> 196
<211> 46
<212> PRT
<213> Homo sapien

<400> 196

Met Lys Ile Tyr Phe Lys Ile Leu Leu Met Phe Leu Lys Lys Tyr Phe
1 5 10 15

Leu Arg Phe His Leu Met Lys Thr Met Lys Tyr Ser Val Phe Tyr Ser
20 25 30

Thr Thr Arg Gln Met Trp Ser Ile Pro Phe Val Phe Phe Phe
35 40 45

<210> 197
<211> 18
<212> PRT
<213> Homo sapien

<400> 197

Met Leu Glu Ala Gly Ile Ser Phe Lys Val Arg Leu Gln Lys Trp Lys
1 5 10 15

Gln Ile

<210> 198
<211> 132
<212> PRT
<213> Homo sapien

<400> 198

Met Phe Tyr Ser Ile Leu Ala Met Leu Arg Asp Arg Gly Ala Leu Gln
1 5 10 15

Asp Leu Met Asn Met Leu Glu Leu Asp Ser Ser Gly His Leu Asp Gly
20 25 30

Pro Gly Gly Ala Ile Leu Lys Lys Leu Gln Gln Asp Ser Asn His Ala
35 40 45

Trp Phe Asn Pro Lys Asp Pro Ile Leu Tyr Leu Leu Glu Ala Ile Met
50 55 60

Val Leu Ser Asp Phe Gln His Asp Leu Leu Ala Cys Ser Met Glu Lys
65 70 75 80

Arg Ile Leu Leu Gln Gln Gln Glu Leu Val Arg Ser Ile Leu Glu Pro
85 90 95

Asn Phe Arg Tyr Pro Trp Ser Ile Pro Phe Thr Leu Lys Pro Glu Leu
100 105 110

Leu Ala Pro Leu Gln Ser Glu Gly Leu Ala Ser Pro Met Ala Ala Gly
115 120 125

Gly Val Trp Pro
130

<210> 199
<211> 226
<212> PRT
<213> Homo sapien

<400> 199

Pro Pro Lys His Leu Lys Ser Lys Phe Gly Gly Met Arg Lys Ala Asp
1 5 10 15

Asp Asp Leu Ile Leu Leu Leu Gly Arg Ile Glu Glu Pro Phe Trp Gln

20 25 30

Asn Phe Lys His Leu Gln Glu Glu Val Phe Gln Lys Ile Lys Thr Leu
35 40 45

Ala Gln Leu Ser Lys Asp Val Gln Asp Val Met Phe Tyr Ser Ile Leu
50 55 60

Ala Met Leu Arg Asp Arg Gly Ala Leu Gln Asp Leu Met Asn Met Leu
65 70 75 80

Glu Leu Asp Ser Ser Gly His Leu Asp Gly Pro Gly Gly Ala Ile Leu
85 90 95

Lys Lys Leu Gln Gln Asp Ser Asn His Ala Trp Phe Asn Pro Lys Asp
100 105 110

Pro Ile Leu Tyr Leu Leu Glu Ala Ile Met Val Leu Ser Asp Phe Gln
115 120 125

His Asp Leu Leu Ala Cys Ser Met Glu Lys Arg Ile Leu Leu Gln Gln
130 135 140

Gln Glu Leu Val Arg Ser Ile Leu Glu Pro Asn Phe Arg Tyr Pro Trp
145 150 155 160

Ser Ile Pro Phe Thr Leu Lys Pro Glu Leu Leu Ala Pro Leu Gln Ser
165 170 175

Glu Gly Leu Ala Ile Thr Tyr Gly Leu Leu Glu Glu Cys Gly Leu Arg
180 185 190

Thr Glu Leu Asp Asn Pro Arg Ser Thr Trp Asp Val Glu Ala Lys Met
195 200 205

Pro Leu Ser Ala Leu Tyr Gly Thr Leu Ser Leu Leu Gln Gln Leu Ala
210 215 220

Glu Ala
225

<210> 200
<211> 37
<212> PRT
<213> Homo sapien

<400> 200

Met Ala Lys His Lys Gly Gly Tyr Gly Lys Tyr Trp Val Thr Leu Ile
1   5   10   15

Ile Gly Leu Asn Ala Thr Asn Asn Ile Ile Ile Val Leu Thr Tyr Phe
            20   25   30

Phe Arg Leu Leu Ser
        35

<210> 201
<211> 28
<212> PRT
<213> Homo sapien

<400> 201

Met Val His Lys Ser Tyr Phe Thr Thr Leu Ser Leu Val Ile Leu Gly
1   5   10   15

Val Trp Pro Cys Lys Ala Ser Ser Gln Arg Phe Cys
            20   25

<210> 202
<211> 77
<212> PRT
<213> Homo sapien

<400> 202

Met Gly Ser Val Cys Val Cys Phe His Arg Ser Thr Thr Ser Glu Val
1   5   10   15

Ser Leu His Leu Cys Ile Phe Thr Ser Gln Gly Gln Gly Pro Gly Asn
            20   25   30

Leu Arg Gly Ser His Ser Phe Ser Leu Pro Gln Thr Met Pro Leu Pro
        35   40   45

Pro Ile Ser Leu Gly Gln Glu Ser Gly Phe Cys Phe Pro Tyr Phe Phe
    50   55   60

Phe Pro Arg His Trp Glu Ala Ser Gly Glu Gln His Gln
65   70   75

<210> 203
<211> 70

<212> PRT
<213> Homo sapien

<400> 203

Met Gly Pro Pro Leu Pro Leu Gly Gly Trp Ser Ser Asp Leu Leu Ala
1 5 10 15

Gln Lys Val Leu Phe Phe His Leu Leu Cys Leu Asn Glu Ser Ser Trp
20 25 30

Thr Tyr Thr Pro Leu Ser Asp Glu Arg Ala Arg Leu Arg Arg Cys Ala
35 40 45

Gly His Leu Leu Arg Ile His Val Gly Ser Ala Ala Pro Gly Gly Gly
50 55 60

Ser Thr Ser Ala Ala Thr
65 70

<210> 204
<211> 37
<212> PRT
<213> Homo sapien

<400> 204

Met Ser Lys Lys Lys Asp Gln Asp Leu Cys Leu Lys Ile Glu Met His
1 5 10 15

Thr Ala Ala Ala Gln Lys Leu Arg Pro Ala Ser Lys Leu His Glu Ala
20 25 30

Leu Val Lys Thr Asp
35

<210> 205
<211> 87
<212> PRT
<213> Homo sapien

<400> 205

Met Pro Ser Val Ala Gln Gly Pro Val Pro Trp His Leu Gly Ser Arg
1 5 10 15

Ser Ala Val Ala Val Phe Glu Phe Leu Val Met Phe Glu Gln Arg Pro
20 25 30

Tyr Val Cys Ile Leu His Trp Ala Pro Gln Ile Thr Trp Pro Ile Leu
35 40 45

Arg Arg Gly Val Ser His Leu Gln Ser Pro Lys Ser Pro Leu Glu Val
50 55 60

Phe Leu Asn Glu Arg Thr Glu Ala Phe Leu Lys Ser Ser Val Gly Glu
65 70 75 80

Thr Val His His His Thr Gln
85

<210> 206
<211> 46
<212> PRT
<213> Homo sapien

<400> 206

Met Ser Pro Gly Thr Ala Met Ala Leu Gly Ala Pro Thr Leu Phe Phe
1 5 10 15

Phe Phe Phe Phe Phe Phe Phe Tyr Asn Gln Pro Ile Arg Asp Leu Ser
20 25 30

Ile Asn Lys Pro Leu Phe Ile Ile Arg Asn Trp Leu Thr Gln
35 40 45

<210> 207
<211> 91
<212> PRT
<213> Homo sapien

<400> 207

Met Ser Ser Pro Gln Ser Ile Glu His Asn His Asp Ser His Glu Leu
1 5 10 15

Pro Thr Pro Pro Ala Ala Ser Ala Gln Arg Glu Ser Pro Leu Gln Val
20 25 30

Cys Leu Ile Ala Glu Pro Ile Phe Phe Leu Pro Gly Gln Gln Leu Leu
35 40 45

Ser Ser Met Ser Arg His Trp Cys Ser Leu Gly Trp Ala Pro Val Thr
50 55 60

Pro Met Glu Ile Leu Asp Gly Cys Tyr Arg Thr Gly Leu Asp Val Arg

65 70 75 80

Gly Leu Gly Asn Gly Ala Gln Glu Ser Ser Ser
85 90

<210> 208
<211> 87
<212> PRT
<213> Homo sapien

<400> 208

Met Cys Val Arg Asn Ser Met Phe Lys Lys Glu Ile Ile Gln Arg Val
1 5 10 15

Thr Asn His Gly Ser Val Gly His Trp Thr Lys Leu Gly Phe Trp Thr
20 25 30

Phe Leu Pro Asn Ile Asn Phe Ala Leu Ala Ser Val Tyr Thr His Thr
35 40 45

His Thr Thr Thr Asn Thr Thr Gln Thr Thr Phe Trp Ala Asn Thr Thr
50 55 60

Arg Arg Gln Arg Arg Leu Pro Gly Leu Lys Leu Gly Ser Leu Pro Ala
65 70 75 80

Pro Gln Phe Ser Gln Gln Leu
85

<210> 209
<211> 55
<212> PRT
<213> Homo sapien

<400> 209

Met Thr Cys Phe Arg Glu Cys Leu Leu Val Tyr Leu Tyr Ser Ile Cys
1 5 10 15

Leu Leu Asn Ser Leu His Lys Leu Glu Leu Leu Ser Arg Arg Leu Arg
20 25 30

Glu Cys Lys Tyr Val Thr His Lys Met His Trp Ser Met Val Asn Lys
35 40 45

Thr Asn His Phe Gly Leu Val
50 55

<210> 210
<211> 58
<212> PRT
<213> Homo sapien

<400> 210

Met Val Ile Phe Tyr Ser Ser Pro Ser Gln Asp Ser Ala Leu Ile Tyr
1 5 10 15

Tyr Ile Pro Phe Ile Leu Leu Tyr Arg Leu Leu Ser Glu Thr His Val
20 25 30

Gln Ile Arg Asp Lys Ile Leu Lys His Ile Thr Pro Ser Leu Val Phe
35 40 45

Ser Ile Gln Ile Leu Arg Asn Ser Cys Tyr
50 55

<210> 211
<211> 37
<212> PRT
<213> Homo sapien

<400> 211

Met Asn Leu Tyr Leu Lys Met Lys Thr Ile Pro Lys Lys Thr Cys Met
1 5 10 15

Ser Lys Thr Glu Leu Phe Leu Pro Phe Thr Pro Lys Tyr Leu Lys Leu
20 25 30

Asn Leu Ser His Phe
35

<210> 212
<211> 99
<212> PRT
<213> Homo sapien

<400> 212

Phe Phe Phe Phe Leu Arg Trp Ser Leu Ala Leu Ser Pro Arg Leu Glu
1 5 10 15

Cys Ser Gly Val Ile Ser Thr His Cys Asn Leu Cys Phe Pro Gly Ser
20 25 30

Ser Asp Ser Arg Ala Ser Pro Thr Phe Gln Val Ala Trp Ile Thr Gly
35 40 45

Val Arg His His Ser Trp Leu Ile Phe Val Leu Leu Val Glu Thr Gly
50 55 60

Phe His His Val Val Gln Ala Val Glu Leu Leu Thr Ser Arg Asp Pro
65 70 75 80

Pro Ala Ser Ala Ser Gln Ser Ala Ala Ile Ile Gly Val Asn His Cys
85 90 95

Ala Arg Pro

<210> 213
<211> 43
<212> PRT
<213> Homo sapien

<400> 213

Met Gln Glu Phe Thr Trp Leu Phe Glu Lys Glu Asn Phe Lys Val Ser
1 5 10 15

Gly Trp Thr Glu Ser His Glu Ala Arg Ser Leu Leu Thr Ala Arg Ser
20 25 30

Leu Glu Lys Gln Val Ser Gly Ser Phe Thr Ser
35 40

<210> 214
<211> 61
<212> PRT
<213> Homo sapien

<400> 214

Met Ala Val Asp Phe Tyr Asn Phe Val Thr Lys Leu Val Val Thr Thr
1 5 10 15

Gly Tyr Leu Arg Ile Ser Phe Leu Ala Tyr Lys Phe Phe Ser Phe Pro
20 25 30

Phe Leu Asp Ser Leu Ser Leu Cys Cys Pro Gly Leu Glu Cys Ser Gly
35 40 45

Val Ile Pro Ala His Tyr Asn Leu Tyr Leu Pro Gly Arg

50 55 60

<210> 215
<211> 127
<212> PRT
<213> Homo sapien

<400> 215

Ser Gln Asn Ile Phe Phe Gly Val Ala Ile Phe Phe Phe Ser Phe Phe
1 5 10 15

Arg Gln Ser Leu Ser Leu Val Ala Gln Ala Arg Val Gln Trp Arg Asp
20 25 30

Pro Gly Ser Leu Gln Pro Leu Pro Pro Gly Phe Lys Arg Phe Leu Gly
35 40 45

Leu Ser Leu Pro Ser Ser Ala Gly Tyr Arg Arg Ala Pro Pro Pro Cys
50 55 60

Pro Ala Leu Leu Tyr Phe Ala Val Glu Thr Gly Phe His His Val Gly
65 70 75 80

Gln Ala Gly Leu Glu Leu Leu Thr Ser Gly Asn Pro Ala Ala Ser Ala
85 90 95

Ser Gln Ser Ala Gly Ile Thr Gly Thr Ser His Cys Thr Gln Pro Tyr
100 105 110

Tyr His Lys Ser Ser Ala Cys Trp Tyr Leu Ile Arg Phe Tyr Leu
115 120 125

<210> 216
<211> 13
<212> PRT
<213> Homo sapien

<400> 216

Met Glu Cys Ser Ser Leu Ala Glu Phe Lys Pro Val Phe
1 5 10

<210> 217
<211> 100
<212> PRT
<213> Homo sapien

<400> 217

Pro Gln Gln Thr Leu Lys Arg Ile Gln Gln Val Leu Ile Lys Cys Cys
1 5 10 15

Leu Ala Phe Tyr Leu Phe Leu Phe Phe Phe Phe Leu Arg Trp Ser Leu
20 25 30

Ala Leu Leu Pro Ser Leu Lys Cys Ser Gly Val Ile Ser Ala His Cys
35 40 45

Asn Leu Arg Leu Pro Gly Leu Gly Asp Ser Leu Ala Ser Ala Ser Arg
50 55 60

Val Ala Gly Met Thr Thr Gly Thr Cys His His Ala Gln Leu Ile Phe
65 70 75 80

Val Phe Leu Val Glu Thr Gly Phe Cys Val Ser Gln Asp Gly Leu Asp
85 90 95

Leu Leu Ile Ser
100

<210> 218
<211> 46
<212> PRT
<213> Homo sapien

<400> 218

Met Glu Gly Gly Glu Met Ser Thr Gln Val Glu Asn Arg Ser Glu Gly
1 5 10 15

Thr Ile Pro Ile Gln Thr Thr Cys Lys Ser His Asn Lys Ala Pro His
20 25 30

Cys Thr Glu Leu Arg His Lys Gln Arg Phe Pro Thr Asp Gly
35 40 45

<210> 219
<211> 72
<212> PRT
<213> Homo sapien

<400> 219

Ile Ser Phe Ile Phe Phe Ser Glu Ala Cys Gln Val Glu Val Arg Leu
1 5 10 15

Leu Leu Ala Tyr Asn Ser Ser Ala Arg Ile Pro Lys Cys Pro Trp Met
20 25 30

Glu Gly Gly Glu Met Ser Pro Gln Val Glu Thr Ser Ile Glu Gly Thr
35 40 45

Ile Pro Phe Ser Lys Pro Val Lys Val Tyr Ile Met Pro Lys Pro Ala
50 55 60

Arg Arg Pro Lys Pro Ala Arg Arg
65 70

<210> 220
<211> 41
<212> PRT
<213> Homo sapien

<400> 220

Met Glu Cys Lys Val Ile Lys Cys Ser Cys Phe His Leu Glu Gly Cys
1 5 10 15

Gly Pro Glu Gly Lys Arg Ser Pro Lys Tyr Pro Pro Pro Trp Cys Ser
20 25 30

Ser Leu Cys Leu Val Pro Ala Arg Ala
35 40

<210> 221
<211> 30
<212> PRT
<213> Homo sapien

<400> 221

Met Asn Ser Phe Gly Tyr Met Thr Pro Ser Lys Phe Phe Lys Lys Glu
1 5 10 15

Ile Thr Phe Lys Thr Thr Tyr Ile Phe Cys Phe Cys Leu Arg
20 25 30

<210> 222
<211> 22
<212> PRT
<213> Homo sapien

<400> 222

Met Leu Gln Ile Gly His Leu Leu Ser Met His Ser Leu Asp Lys Asn
1 5 10 15

Ile Gly Gln Val Gly Met
20

<210> 223
<211> 18
<212> PRT
<213> Homo sapien

<400> 223

Met Ser Asp Arg Val Val Ala Leu Leu Glu Val Phe Phe Pro Phe Gln
1 5 10 15

Arg Glu

<210> 224
<211> 133
<212> PRT
<213> Homo sapien

<400> 224

Met Gly Asn Ser Ile Asp Thr Val Arg Tyr Gly Lys Glu Ser Asp Leu
1 5 10 15

Gly Asp Val Ser Glu Glu His Gly Glu Trp Asn Lys Glu Ser Ser Asn
20 25 30

Asn Glu Gln Asp Asn Ser Leu Leu Glu Gln Tyr Leu Thr Ser Val Gln
35 40 45

Gln Leu Glu Asp Ala Asp Glu Arg Thr Asn Phe Asp Thr Glu Thr Arg
50 55 60

Asp Ser Lys Leu His Ile Ala Cys Phe Pro Val Gln Leu Asp Thr Leu
65 70 75 80

Ser Asp Gly Ala Ser Val Asp Glu Ser His Gly Ile Ser Pro Pro Leu
85 90 95

Gln Gly Glu Ile Ser Gln Thr Gln Glu Asn Ser Lys Leu Asn Ala Glu
100 105 110

Val Gln Gly Gln Gln Pro Glu Cys Asp Ser Thr Phe Gln Leu Leu His
115 120 125

Val Gly Val Thr Val
    130

<210> 225
<211> 50
<212> PRT
<213> Homo sapien

<400> 225

Met Arg Asn Ser Ser Pro Ile Leu Thr Pro Ala Leu Phe Ser Phe His
1   5   10   15

Met Tyr Ile Gly Pro Leu Ile Arg Ile Phe Lys Lys Phe Pro Arg Pro
            20   25   30

Pro Asn Leu Thr Ile Asp Asp Pro Leu Ser Leu Phe Arg Arg Asn Tyr
        35   40   45

Ile Gly
    50

<210> 226
<211> 43
<212> PRT
<213> Homo sapien

<400> 226

Met His Ser Phe Phe Leu Ser Met Leu Cys Pro Glu Ala Leu Arg Val
1   5   10   15

Leu Leu Lys Gln Ala Ala Gly Leu Leu Arg Glu Ile Lys Gly Phe Ile
            20   25   30

Ser Thr Thr Arg Cys Gln Asn Leu His Phe Glu
        35   40

<210> 227
<211> 99
<212> PRT
<213> Homo sapien

<400> 227

Met Leu Glu Arg Arg Ser Val Met Asp Arg Arg Arg Ala Gly Asn Ser
1   5   10   15

Pro Pro Arg Ile Glu Lys Cys Leu Leu Gly Arg Glu Glu Gly Glu Ala

20 25 30

Gly Ala Gly Pro Ser Pro Gly Ser Leu Leu Gly Pro Gln Lys Ala Leu
35 40 45

Asn Gln Ala Pro Ser Leu Gln Gly Lys Pro Arg Pro Gln Pro Asp Asn
50 55 60

Leu Glu Gly Arg Lys Ser Gln Thr Leu Gly Leu Phe Phe Gly Gly Ile
65 70 75 80

Ile Gly Phe Phe Phe Phe Met Phe Leu Leu Glu Phe Cys Leu Leu Ala
85 90 95

Asn Ser Val

<210> 228
<211> 44
<212> PRT
<213> Homo sapien

<400> 228

Met Lys Ser Ile Gln Leu Lys Phe Ser Tyr Ile Ile Glu Pro Gln Leu
1 5 10 15

Asn Gly Met Asn Gly Ile Gly Asn Leu Leu Glu Met Ile Phe Met Ile
20 25 30

Thr Phe Val Val Ile Pro Phe Ser Trp Leu Arg Phe
35 40

<210> 229
<211> 41
<212> PRT
<213> Homo sapien

<400> 229

Tyr Phe Pro Leu Gln Ile Trp Ile Ser Glu Asp Ser Asn Asn Ile Glu
1 5 10 15

Ala Val Asn Gln Trp Lys Glu Thr Val Ile Asn Pro Glu Lys Val Val
20 25 30

Ile Arg Trp His Lys Leu Asn Pro Ser
35 40

<210> 230
<211> 48
<212> PRT
<213> Homo sapien

<400> 230

Met Leu Lys Gly His Tyr Gln Tyr Gly Met Glu Asp Leu Ser Phe His
1 5 10 15

Thr Phe Ser Ser Ser Phe Leu Asn Phe Leu Leu Leu Phe Leu Leu Ser
20 25 30

Cys Met Val Ala Pro Phe Pro Phe Leu Leu Ser Val Pro Ser Lys Gln
35 40 45

<210> 231
<211> 108
<212> PRT
<213> Homo sapien

<400> 231

Phe Leu Lys Arg Gln Ser Ile Ser Leu Leu Pro Gln Leu Glu Cys Ser
1 5 10 15

Gly Thr Ile Ile Val His His Thr Leu Glu Leu Leu Gly Lys Gly Ser
20 25 30

Ser Leu Ala Ser Ala Ser Gln Val Ala Arg Tyr Thr Gly Met Cys Tyr
35 40 45

His Ala Trp Leu Ile Lys Lys Ile Phe Leu Glu Met Arg Ser Cys Cys
50 55 60

Val Ala Gln Ala Gly Leu Lys Leu Leu Gly Ser Asn Asn Pro Pro Thr
65 70 75 80

Leu Ala Ser Gln Ser Ala Gly Ile Thr Gly Val Ser His Ser Thr Ala
85 90 95

Pro Tyr Leu Gln Ile Leu Asn Gln Ala Ile Ala Ile
100 105

<210> 232
<211> 64
<212> PRT

<213> Homo sapien

<400> 232

Met Ser Pro Arg Ala Pro Phe Ala Pro Gly Cys Pro Gln Pro Leu Val
1 5 10 15

Val Phe Tyr Val Cys Phe Phe Phe Phe Leu Ile Phe Cys Phe Val Lys
20 25 30

Lys His His Tyr Met Phe Leu Tyr Pro Arg Leu Lys Thr Phe Gly Asn
35 40 45

Leu Ile Ser Asn Ile Lys Ile Gln Ile Lys Thr His Ser Thr Ile Pro
50 55 60

<210> 233
<211> 35
<212> PRT
<213> Homo sapien

<400> 233

Met Cys Val Asn Ala Ser Thr Val Gly Gln Met Cys Glu Asn Glu Leu
1 5 10 15

Lys His Met Leu Arg Ile Lys Val Asn Arg Arg Asn Phe Glu Arg Phe
20 25 30

Pro Leu Met
35

<210> 234
<211> 72
<212> PRT
<213> Homo sapien

<400> 234

Met Asn Ile Phe Pro Trp Ala Gly Gly Pro Trp Ser Leu Pro Gln Ala
1 5 10 15

Arg Tyr Arg Ala Pro Ala Cys Ala Pro Thr Asn His Gly Lys Gln Arg
20 25 30

Arg Pro Pro His Leu Lys Ser Trp Pro Val Val Val Ser Ser Val Phe
35 40 45

Leu Leu Ser Glu Gln Asn Val Leu Lys Leu Glu Leu Thr Lys Val Lys

50 55 60

Ser Ser Lys Thr Thr Tyr Ala Thr
65 70

<210> 235
<211> 1163
<212> PRT
<213> Homo sapien

<400> 235

Met Asp Arg Asn Arg Glu Ala Glu Met Glu Leu Arg Arg Gly Pro Ser
1 5 10 15

Pro Thr Arg Ala Gly Arg Gly His Glu Val Asp Gly Asp Lys Ala Thr
20 25 30

Cys His Thr Cys Cys Ile Cys Gly Lys Ser Phe Pro Phe Gln Ser Ser
35 40 45

Leu Ser Gln His Met Arg Lys His Thr Gly Glu Lys Pro Tyr Lys Cys
50 55 60

Pro Tyr Cys Asp His Arg Ala Ser Gln Lys Gly Asn Leu Lys Ile His
65 70 75 80

Ile Arg Ser His Arg Thr Gly Thr Leu Ile Gln Gly His Glu Pro Glu
85 90 95

Ala Gly Glu Ala Pro Leu Gly Glu Met Arg Ala Ser Glu Gly Leu Asp
100 105 110

Ala Cys Ala Ser Pro Thr Lys Ser Ala Ser Ala Cys Asn Arg Leu Leu
115 120 125

Asn Gly Ala Ser Gln Ala Asp Gly Ala Arg Val Leu Asn Gly Ala Ser
130 135 140

Gln Ala Asp Ser Gly Arg Val Leu Leu Arg Ser Ser Lys Lys Gly Ala
145 150 155 160

Glu Gly Ser Ala Cys Ala Pro Gly Glu Ala Lys Ala Ala Val Gln Cys
165 170 175

Ser Phe Cys Lys Ser Gln Phe Glu Arg Lys Lys Asp Leu Glu Leu His

180 185 190

Val His Gln Ala His Lys Pro Phe Lys Cys Arg Leu Cys Ser Tyr Ala
195 200 205

Thr Leu Arg Glu Glu Ser Leu Leu Ser His Ile Glu Arg Asp His Ile
210 215 220

Thr Ala Gln Gly Pro Gly Ser Gly Glu Ala Cys Val Glu Asn Gly Lys
225 230 235 240

Pro Glu Leu Ser Pro Gly Glu Phe Pro Cys Glu Val Cys Gly Gln Ala
245 250 255

Phe Ser Gln Thr Trp Phe Leu Lys Ala His Met Lys Lys His Arg Gly
260 265 270

Ser Phe Asp His Gly Cys His Ile Cys Gly Arg Arg Phe Lys Glu Pro
275 280 285

Trp Phe Leu Lys Asn His Met Lys Ala His Gly Pro Lys Thr Gly Ser
290 295 300

Lys Asn Arg Pro Lys Ser Glu Leu Asp Pro Ile Ala Thr Ile Asn Asn
305 310 315 320

Val Val Gln Glu Glu Val Ile Val Ala Gly Leu Ser Leu Tyr Glu Val
325 330 335

Cys Ala Lys Cys Gly Asn Leu Phe Thr Asn Leu Asp Ser Leu Asn Ala
340 345 350

His Asn Ala Ile His Arg Arg Val Glu Ala Ser Arg Thr Arg Ala Pro
355 360 365

Ala Glu Glu Gly Ala Glu Gly Pro Ser Asp Thr Lys Gln Phe Phe Leu
370 375 380

Gln Cys Leu Asn Leu Arg Pro Ser Ala Ala Gly Asp Ser Cys Pro Gly
385 390 395 400

Thr Gln Ala Gly Arg Arg Val Ala Glu Leu Asp Pro Val Asn Ser Tyr
405 410 415

```
Gln Ala Trp Gln Leu Ala Thr Arg Gly Lys Val Ala Glu Pro Ala Glu
            420                 425                 430

Tyr Leu Lys Tyr Gly Ala Trp Asp Glu Ala Leu Ala Gly Asp Val Ala
        435                 440                 445

Phe Asp Lys Asp Arg Arg Glu Tyr Val Leu Val Ser Gln Glu Lys Arg
    450                 455                 460

Lys Arg Glu Gln Asp Ala Pro Ala Ala Gln Gly Pro Pro Arg Lys Arg
465                 470                 475                 480

Ala Ser Gly Pro Gly Asp Pro Ala Pro Ala Gly His Leu Asp Pro Arg
                485                 490                 495

Ser Ala Ala Arg Pro Asn Arg Arg Ala Ala Ala Thr Thr Gly Gln Gly
            500                 505                 510

Lys Ser Ser Glu Cys Phe Glu Cys Gly Lys Ile Phe Arg Thr Tyr His
        515                 520                 525

Gln Met Val Leu His Ser Arg Val His Arg Arg Ala Arg Arg Glu Arg
    530                 535                 540

Asp Ser Asp Gly Asp Arg Ala Ala Arg Ala Arg Cys Gly Ser Leu Ser
545                 550                 555                 560

Glu Gly Asp Ser Ala Ser Gln Pro Ser Ser Pro Gly Ser Ala Cys Ala
                565                 570                 575

Ala Ala Asp Ser Pro Gly Ser Gly Leu Ala Asp Glu Ala Ala Glu Asp
            580                 585                 590

Ser Gly Glu Glu Gly Ala Pro Glu Pro Ala Pro Gly Gly Gln Pro Arg
        595                 600                 605

Arg Cys Cys Phe Ser Glu Glu Val Thr Ser Thr Glu Leu Ser Ser Gly
    610                 615                 620

Asp Gln Ser His Lys Met Gly Asp Asn Ala Ser Glu Arg Asp Thr Gly
625                 630                 635                 640

Glu Ser Lys Ala Gly Ile Ala Ala Ser Val Ser Ile Leu Glu Asn Ser
                645                 650                 655
```

Ser Arg Glu Thr Ser Arg Arg Gln Glu Gln His Arg Phe Ser Met Asp
660 665 670

Leu Lys Met Pro Ala Phe His Pro Lys Gln Glu Val Pro Val Pro Gly
675 680 685

Asp Gly Val Glu Phe Pro Ser Ser Thr Gly Ala Glu Gly Gln Thr Gly
690 695 700

His Pro Ala Glu Lys Leu Ser Asp Leu His Asn Lys Glu His Ser Gly
705 710 715 720

Gly Gly Lys Arg Ala Leu Ala Pro Asp Leu Met Pro Leu Asp Leu Ser
725 730 735

Ala Arg Ser Thr Arg Asp Asp Pro Ser Asn Lys Glu Thr Ala Ser Ser
740 745 750

Leu Gln Ala Ala Leu Val Val His Pro Cys Pro Tyr Cys Ser His Lys
755 760 765

Thr Tyr Tyr Pro Glu Val Leu Trp Met His Lys Arg Ile Trp His Arg
770 775 780

Val Ser Cys Asn Ser Val Ala Pro Pro Trp Ile Gln Pro Asn Gly Tyr
785 790 795 800

Lys Ser Ile Arg Ser Asn Leu Val Phe Leu Ser Arg Ser Gly Arg Thr
805 810 815

Gly Pro Pro Pro Ala Leu Gly Gly Lys Glu Cys Gln Pro Leu Leu Leu
820 825 830

Ala Arg Phe Thr Arg Thr Gln Val Pro Gly Gly Met Pro Gly Ser Lys
835 840 845

Ser Gly Ser Ser Pro Leu Gly Val Val Thr Lys Ala Ala Ser Met Pro
850 855 860

Lys Asn Lys Glu Ser His Ser Gly Gly Pro Cys Ala Leu Trp Ala Pro
865 870 875 880

Gly Pro Asp Gly Tyr Arg Gln Thr Lys Pro Cys His Gly Gln Glu Pro
885 890 895

His Gly Ala Ala Thr Gln Gly Pro Leu Ala Lys Pro Arg Gln Glu Ala
900 905 910

Ser Ser Lys Pro Val Pro Ala Pro Gly Gly Gly Gly Phe Ser Arg Ser
915 920 925

Ala Thr Pro Thr Pro Thr Val Ile Ala Arg Ala Gly Ala Gln Pro Ser
930 935 940

Ala Asn Ser Lys Pro Val Glu Lys Phe Gly Val Pro Pro Ala Gly Ala
945 950 955 960

Gly Phe Ala Pro Thr Asn Lys His Ser Ala Pro Asp Ser Leu Lys Ala
965 970 975

Lys Phe Ser Ala Gln Pro Gln Gly Pro Pro Pro Ala Lys Gly Glu Gly
980 985 990

Gly Ala Pro Pro Leu Pro Pro Arg Glu Pro Pro Ser Lys Ala Ala Gln
995 1000 1005

Glu Leu Arg Thr Leu Ala Thr Cys Ala Ala Gly Ser Arg Gly Asp
1010 1015 1020

Ala Ala Leu Gln Ala Gln Pro Gly Val Ala Gly Ala Pro Pro Val
1025 1030 1035

Leu His Ser Ile Lys Gln Glu Pro Val Ala Glu Gly His Glu Lys
1040 1045 1050

Arg Leu Asp Ile Leu Asn Ile Phe Lys Thr Tyr Ile Pro Lys Asp
1055 1060 1065

Phe Ala Thr Leu Tyr Gln Gly Trp Gly Val Ser Gly Pro Gly Leu
1070 1075 1080

Glu His Arg Gly Thr Leu Arg Thr Gln Ala Arg Pro Gly Glu Phe
1085 1090 1095

Val Cys Ile Glu Cys Gly Lys Ser Phe His Gln Pro Gly His Leu
1100 1105 1110

Arg Ala His Met Arg Ala His Ser Val Val Phe Glu Ser Asp Gly

1115 1120 1125

Pro Arg Gly Ser Glu Val His Thr Thr Ser Ala Asp Ala Pro Lys
1130 1135 1140

Gln Gly Arg Asp His Ser Asn Thr Gly Thr Val Gln Thr Val Pro
1145 1150 1155

Leu Arg Lys Gly Thr
1160

<210> 236
<211> 55
<212> PRT
<213> Homo sapien

<400> 236

Met Cys Val Phe Cys Gly Phe Phe Cys Ser Arg Phe Val Arg Glu Met
1 5 10 15

Trp Gly Asn Phe Gly Pro Lys Thr Asn Phe Thr Pro Gly Thr Pro Phe
20 25 30

Cys Pro Trp Leu Ser Pro Asn Leu Phe Cys Leu Val Val Val Trp Phe
35 40 45

Tyr Arg Leu Leu Ile Phe Tyr
50 55

<210> 237
<211> 156
<212> PRT
<213> Homo sapien

<400> 237

Met Pro Met Glu Gly His Thr Leu Cys Met Arg Ile Arg Gly Ser Trp
1 5 10 15

Leu Ala Ala Arg Leu Pro Val Met Pro Phe Glu Gly Asp Val Gly Pro
20 25 30

Trp Val Arg Met Lys Val Phe Ile Cys His Ser Ser Ser Pro Gln Val
35 40 45

Ala Ile His Leu Gly Gly Gly Arg Glu Gly Ser Ala Leu Ala Ile Val
50 55 60

Tyr Pro Ala Ser Leu Arg Phe Ile Asp Leu His Lys Arg Leu Cys Ser
65 70 75 80

Gly Lys Gly Arg Gly Pro Gln Lys Gly Ala Trp Gln Asp Arg Trp Met
85 90 95

Leu Tyr Gly His Met Glu Ile Thr Pro Ser Ser Leu Ala Pro Ala Ser
100 105 110

Ala Ser Arg Pro Leu His Gly Val Arg Cys Phe Cys Ala Cys Cys Pro
115 120 125

Thr Ser Leu His Ser Arg Ala Leu Ile Asn His Phe Asp Pro Pro Leu
130 135 140

Ala Glu Gly Ser Pro Leu Tyr Arg Val Gln Ser Leu
145 150 155

<210> 238
<211> 86
<212> PRT
<213> Homo sapien

<400> 238

Met Met Asn Phe Leu Cys Leu Asn Phe Arg Asp Ile Trp Cys Asp Phe
1 5 10 15

His Leu Tyr Leu Met Leu Pro Leu Leu Pro Ser Leu Leu Asn Thr Ser
20 25 30

Lys Asn Ser Glu His Ile Leu Ile Pro Pro Val Phe Tyr Phe Tyr Asp
35 40 45

Leu Asp Ile Leu His His Lys Ile Pro Pro Asn Trp Asp Tyr Val Phe
50 55 60

Glu Val Ile His Phe Thr Ile Ile Thr Thr Ile Thr Ile Ile Phe Ile
65 70 75 80

Val Cys Phe Val Pro Gly
85

<210> 239
<211> 289

<212> PRT
<213> Homo sapien

<400> 239

Ala Asp Leu Ser Phe Ile Glu Asp Thr Val Ala Phe Pro Glu Lys Glu
1 5 10 15

Glu Asp Glu Glu Glu Glu Glu Glu Gly Val Glu Trp Gly Tyr Glu Glu
20 25 30

Gly Val Glu Trp Gly Leu Val Phe Pro Asp Ala Asn Gly Glu Tyr Gln
35 40 45

Ser Pro Ile Asn Leu Asn Ser Arg Glu Ala Arg Tyr Asp Pro Ser Leu
50 55 60

Leu Asp Val Arg Leu Ser Pro Asn Tyr Val Val Cys Arg Asp Cys Glu
65 70 75 80

Val Thr Asn Asp Gly His Thr Ile Gln Val Ile Leu Lys Ser Lys Ser
85 90 95

Val Leu Ser Gly Gly Pro Leu Pro Gln Gly His Glu Phe Glu Leu Tyr
100 105 110

Glu Val Arg Phe His Trp Gly Arg Glu Asn Gln Arg Gly Ser Glu His
115 120 125

Thr Val Asn Phe Lys Ala Phe Pro Met Glu Leu His Leu Ile His Trp
130 135 140

Asn Ser Thr Leu Phe Gly Ser Ile Asp Glu Ala Val Gly Lys Pro His
145 150 155 160

Gly Ile Ala Ile Ile Ala Leu Phe Val Gln Ile Gly Lys Glu His Val
165 170 175

Gly Leu Lys Ala Val Thr Glu Ile Leu Gln Asp Ile Gln Tyr Lys Gly
180 185 190

Lys Ser Lys Thr Ile Pro Cys Phe Asn Pro Asn Thr Leu Leu Pro Asp
195 200 205

Pro Leu Leu Arg Asp Tyr Trp Val Tyr Glu Gly Ser Leu Thr Ile Pro
210 215 220

Pro Cys Ser Glu Gly Val Thr Trp Ile Leu Phe Arg Tyr Pro Leu Thr
225 230 235 240

Ile Ser Gln Leu Gln Ile Glu Glu Phe Arg Arg Leu Arg Thr His Val
245 250 255

Lys Gly Ala Glu Leu Val Glu Gly Cys Asp Gly Ile Leu Gly Asp Asn
260 265 270

Phe Arg Pro Thr Gln Pro Leu Ser Asp Arg Val Ile Arg Ala Ala Phe
275 280 285

Gln

<210> 240
<211> 59
<212> PRT
<213> Homo sapien

<400> 240

Met Cys Gln Ile Asp Arg Gln Asp Leu Val Leu Leu Lys Leu Val Ile
1 5 10 15

Tyr Cys Ser Arg His Leu Lys Gly Trp Arg Arg Ser Glu His Tyr Val
20 25 30

Pro Ala Arg Ala Ser Ile Thr Leu Arg Arg Ser Thr Ser His Leu Val
35 40 45

Ala Arg Ser Pro Asn Met Ser Ser Ser Gly Val
50 55

<210> 241
<211> 41
<212> PRT
<213> Homo sapien

<400> 241

Met Leu Leu Asn Gly Leu His Asn Pro Ala Leu Lys His Leu Arg Asp
1 5 10 15

Leu Cys Lys Thr Phe Pro Trp Ser Leu Cys Phe Ser His Ile Asn Gln
20 25 30

Leu Ala Tyr Phe Ser His Ser Pro Ser
35 40

<210> 242
<211> 80
<212> PRT
<213> Homo sapien

<400> 242

Met Asn Cys Leu Tyr Pro Ser Pro Met Cys Phe Tyr Arg Ser Cys Leu
1 5 10 15

Val His Phe Val Ala Asp Leu Leu Gly Asp Phe Thr Glu Gly Lys Val
20 25 30

Ser Ser Lys Leu Tyr Asp Asp Phe Met Leu Ile Asp Leu Leu Ser Ser
35 40 45

Gly Ser Trp Glu Thr His Ser Ala Ile Ser Leu Leu Ser Tyr Phe Ser
50 55 60

Tyr Asp Ala Gln Pro Pro Lys Ala Thr Arg Glu Gln Tyr Arg Val Pro
65 70 75 80

<210> 243
<211> 45
<212> PRT
<213> Homo sapien

<400> 243

Glu Arg Pro Gly Met Leu Asp Phe Thr Gly Lys Ala Lys Trp Asp Ala
1 5 10 15

Trp Asn Glu Leu Lys Gly Thr Ser Lys Glu Asp Ala Met Lys Ala Tyr
20 25 30

Ile Asn Lys Val Glu Glu Leu Lys Lys Lys Tyr Gly Ile
35 40 45

<210> 244
<211> 24
<212> PRT
<213> Homo sapien

<400> 244

Met Cys Leu Asn Phe Ser Phe Asn Tyr Leu Ile Pro Phe Ala Gln Glu

1 5 10 15

Ile Thr Ile Ser Leu Phe Phe Phe
20

<210> 245
<211> 69
<212> PRT
<213> Homo sapien

<400> 245

Leu Phe Phe Gln Leu Phe Asp Thr Phe Cys Pro Arg Asp Tyr Tyr Leu
1 5 10 15

Ser Leu Phe Phe Phe Ser Phe Lys Thr Glu Cys Cys Ser Val Thr Gln
20 25 30

Val Gly Val Gln Trp His Asn Ser Ala Ser Leu Gln Pro Leu Pro Pro
35 40 45

Arg Leu Lys Arg Ser Ser His Leu Ser Leu Pro Ser Ser Trp Asp His
50 55 60

Arg His Ile Pro Pro
65

<210> 246
<211> 39
<212> PRT
<213> Homo sapien

<400> 246

Met Glu Thr Lys His His Ser His Lys Lys Ser Asn Ser Ile Leu Asn
1 5 10 15

His Trp Lys Val Thr Ile Pro Leu Tyr Ser Phe Pro Lys Leu Phe Val
20 25 30

Ala Lys Ser Tyr Arg Lys Glu
35

<210> 247
<211> 93
<212> PRT
<213> Homo sapien

<400> 247

Leu Leu Gln Ala Leu Lys Lys Ile Phe Phe Leu Asn Ser Leu Thr Leu
1   5   10   15

Ser Pro Arg Leu Glu Ala Ser Asn Val Ile Ser Ala His Cys Asn Leu
            20   25   30

His Ser Arg Val Ala Gly Ile Thr Asp Met His His His Pro Gln Leu
        35   40   45

Ile Phe Val Phe Leu Val Glu Thr Gly Phe Arg His Val Gly Gln Ala
    50   55   60

Gly Leu Ala Leu Leu Ala Leu Arg Asp Pro Pro Pro Leu Ala Phe Gln
65   70   75   80

Ser Ala Gly Ile Thr Gly Val Ser His Cys Thr Trp Pro
                85   90

<210> 248
<211> 51
<212> PRT
<213> Homo sapien

<400> 248

Met Phe Phe Phe Phe Val Phe Phe Phe Phe Leu Phe Ala Arg Phe Ser
1   5   10   15

Arg Asn Val Gly Asp Leu Trp Ala Gly Lys Pro Phe Pro Pro Gly His
            20   25   30

Val Leu Pro Arg Tyr Pro His Leu Phe Phe Phe Phe Phe Phe Phe Cys
        35   40   45

Phe Ile Thr
    50

<210> 249
<211> 62
<212> PRT
<213> Homo sapien

<400> 249

Met Asn Phe Thr Leu Ala Ile Phe His Tyr Phe Ser Leu Ser Gln Met
1   5   10   15

Ser Val Leu Met Arg Gln Leu Ala Leu Thr Gly Ala Thr Leu Met Cys
20 25 30

His Leu Pro Thr Phe Asn Phe Trp Val Lys Ala Glu Arg Glu Lys Leu
35 40 45

Met Asp Phe Ser Phe Ser Arg Arg Asp Lys Asn Gln Leu His
50 55 60

<210> 250
<211> 190
<212> PRT
<213> Homo sapien

<400> 250

Met Lys Leu Gln Leu Arg Ile Lys Ser Leu Thr Gln Asn Arg Thr Thr
1 5 10 15

Thr Trp Lys Leu Asn Asn Leu Leu Leu Asn Asp Tyr Trp Val Asn Lys
20 25 30

Lys Ile Lys Ala Glu Ile Asn Lys Phe Phe Glu Thr Ile Glu Asn Lys
35 40 45

Asp Thr Met Tyr Gln Asn Thr Ala Lys Ala Val Phe Arg Gly Lys Phe
50 55 60

Ile Ala Leu Asn Thr His Ile Arg Asn Trp Glu Ile Pro Lys Ile Asn
65 70 75 80

Val Leu Thr Ser Gln Leu Lys Glu Leu Glu Lys Arg Glu Gln Thr His
85 90 95

Ser Lys Gln Glu Ile Thr Lys Ile Ile Ala Glu Leu Lys Glu Ile Glu
100 105 110

Thr Gln Lys Ala Leu Gln Lys Ile Ser Asp Ser Arg Ser Trp Phe Phe
115 120 125

Glu Lys Ile Asn Lys Thr Asp Arg Leu Leu Ala Arg Ile Ile Lys Lys
130 135 140

Lys Arg Glu Lys Asn Gln Ile Asp Thr Ile Lys Asn Asp Lys Gly Asp
145 150 155 160

Ile Thr Thr Asn Pro Thr Glu Ile Gln Thr Ala Ile Arg Glu Cys Tyr
165 170 175

Gln His Leu Tyr Ile Asn Lys Leu Glu Asn Leu Glu Glu Ile
180 185 190

<210> 251
<211> 132
<212> PRT
<213> Homo sapien

<400> 251

Met Pro Val Leu Ser Pro Pro Leu His Met Pro Tyr Pro Ala Ala Lys
1 5 10 15

Leu Asp Ser Val Leu Pro Asp Lys Thr Trp Tyr Trp His Leu Tyr Ala
20 25 30

Ser Val Cys Leu Pro Ser Thr Phe Lys Lys Pro Leu Gln Ser Ala Asp
35 40 45

Thr Lys Lys Gln Ser His Thr Cys Ser Lys Ser Ala Cys Phe Pro Leu
50 55 60

Ile Ser Ala Ser Cys Gln Arg His Cys Leu Thr Ser Ser Ser Leu Leu
65 70 75 80

Ser Ile Cys Val Pro His Lys Thr Leu Arg Asp Ser Ala Ser Tyr Val
85 90 95

Tyr Gly Leu Trp Val Phe Ile Ser Thr Val Pro Cys Leu Thr Leu Ser
100 105 110

Pro Cys Gly Glu Tyr Thr His Pro Thr Pro Thr Val Pro Cys Thr Ser
115 120 125

Val Ala Ala Gln
130

<210> 252
<211> 30
<212> PRT
<213> Homo sapien

<400> 252

Met Gln Phe Arg Ile His Ala Ser Phe Ser Val Lys Trp Arg Ser Tyr

1 5 10 15

Ser Phe Asn Ser Glu Asn Ser Gln Leu Asn Lys Gln Pro Leu
20 25 30

<210> 253
<211> 49
<212> PRT
<213> Homo sapien

<400> 253

Met Arg Val Val Trp Gly Trp Arg Cys Gly Cys Val Gly Val Leu Val
1 5 10 15

Leu Val Val Gly Gly Cys Val Glu Trp Ala Val Val Phe Gly Val Cys
20 25 30

Val Gly Cys Val Val Trp Val Gly Arg Trp Trp Cys Asp Val Val Val
35 40 45

Trp

<210> 254
<211> 54
<212> PRT
<213> Homo sapien

<400> 254

Met Lys Lys Ser Val Ser Cys Cys Ser Ser Leu Trp Val Ser Leu Ser
1 5 10 15

Lys Asp Glu Asn Ala Glu Val Gly Arg Gly Asp Ser Leu Leu Gly Thr
20 25 30

Gly Arg Cys Gly Leu Pro Ile Thr Arg Leu Lys Leu Thr Ser Leu Pro
35 40 45

Ser Ser Pro Thr Val Val
50

<210> 255
<211> 1088
<212> PRT
<213> Homo sapien

<400> 255

```
Asp Asp Ser Leu Ile Ser Ser Ala Thr Ala Ile Met Glu Ala Val Val
1               5                   10                  15

Arg Glu Trp Ile Leu Leu Glu Lys Gly Ser Ile Glu Ser Leu Arg Thr
            20                  25                  30

Phe Leu Leu Thr Tyr Val Leu Gln Arg Pro Asn Leu Gln Lys Tyr Val
        35                  40                  45

Arg Glu Gln Ile Leu Leu Ala Val Ala Val Ile Val Lys Arg Gly Ser
    50                  55                  60

Leu Asp Lys Ser Ile Asp Cys Lys Ser Ile Phe His Glu Val Ser Gln
65                  70                  75                  80

Leu Ile Ser Ser Gly Asn Pro Thr Val Gln Thr Leu Ala Cys Ser Ile
                85                  90                  95

Leu Thr Ala Leu Leu Ser Glu Phe Ser Ser Ser Ser Lys Thr Ser Asn
            100                 105                 110

Ile Gly Leu Ser Met Glu Phe His Gly Asn Cys Lys Arg Val Phe Gln
        115                 120                 125

Glu Glu Asp Leu Arg Gln Ile Phe Met Leu Thr Val Glu Val Leu Gln
    130                 135                 140

Glu Phe Ser Arg Arg Glu Asn Leu Asn Ala Gln Met Ser Ser Val Phe
145                 150                 155                 160

Gln Arg Tyr Leu Ala Leu Ala Asn Gln Val Leu Ser Trp Asn Phe Leu
                165                 170                 175

Pro Pro Asn Leu Gly Arg His Tyr Ile Ala Met Phe Glu Ser Ser Gln
            180                 185                 190

Asn Val Leu Leu Lys Pro Thr Glu Ser Leu Arg Glu Thr Leu Leu Asp
        195                 200                 205

Ser Arg Val Met Glu Leu Phe Phe Thr Val His Arg Lys Ile Arg Glu
    210                 215                 220

His Ser Asp Met Ala Gln Asp Ser Leu Gln Cys Leu Ala Gln Leu Ala
225                 230                 235                 240
```

Ser Leu His Gly Pro Ile Phe Pro Asp Glu Gly Ser Gln Val Asp Tyr
245 250 255

Leu Ala His Phe Ile Glu Gly Leu Leu Asn Thr Ile Asn Gly Ile Glu
260 265 270

Ile Glu Asp Ser Glu Ala Val Gly Ile Ser Ser Ile Ile Ser Asn Leu
275 280 285

Ile Thr Val Phe Pro Arg Asn Val Leu Thr Ala Ile Pro Ser Glu Leu
290 295 300

Phe Ser Ser Phe Val Asn Cys Leu Thr His Leu Thr Cys Ser Phe Gly
305 310 315 320

Arg Ser Ala Ala Leu Glu Glu Val Leu Asp Lys Asp Asp Met Val Tyr
325 330 335

Met Glu Ala Tyr Asp Lys Leu Leu Glu Ser Trp Leu Thr Leu Val Gln
340 345 350

Asp Asp Lys His Phe His Lys Gly Phe Phe Thr Gln His Ala Val Gln
355 360 365

Val Phe Asn Ser Tyr Ile Gln Cys His Leu Ala Ala Pro Asp Gly Thr
370 375 380

Arg Asn Leu Thr Ala Asn Gly Val Ala Ser Arg Glu Glu Glu Glu Ile
385 390 395 400

Ser Glu Leu Gln Glu Asp Asp Arg Asp Gln Phe Ser Asp Gln Leu Ala
405 410 415

Ser Val Gly Met Leu Gly Arg Ile Ala Ala Glu His Cys Ile Pro Leu
420 425 430

Leu Thr Ser Leu Leu Glu Glu Arg Val Thr Arg Leu His Gly Gln Leu
435 440 445

Gln Arg His Gln Gln Gln Leu Leu Ala Ser Pro Gly Ser Ser Thr Val
450 455 460

Asp Asn Lys Met Leu Asp Asp Leu Tyr Glu Asp Ile His Trp Leu Ile

465 470 475 480

Leu Val Thr Gly Tyr Leu Leu Ala Asp Asp Thr Gln Gly Glu Thr Pro
485 490 495

Leu Ile Pro Pro Glu Ile Met Glu Tyr Ser Ile Lys His Ser Ser Glu
500 505 510

Val Asp Ile Asn Thr Thr Leu Gln Ile Leu Gly Ser Pro Gly Glu Lys
515 520 525

Ala Ser Ser Ile Pro Gly Tyr Asn Arg Thr Asp Ser Val Ile Arg Leu
530 535 540

Leu Ser Ala Ile Leu Arg Val Ser Glu Val Glu Ser Arg Ala Ile Arg
545 550 555 560

Ala Asp Leu Thr His Leu Leu Ser Pro Gln Met Gly Lys Asp Ile Val
565 570 575

Trp Phe Leu Lys Arg Trp Ala Lys Thr Tyr Leu Leu Val Asp Glu Lys
580 585 590

Leu Tyr Asp Gln Ile Ser Leu Pro Phe Ser Thr Ala Phe Gly Ala Asp
595 600 605

Thr Glu Gly Ser Gln Trp Ile Ile Gly Tyr Leu Leu Gln Lys Val Ile
610 615 620

Ser Asn Leu Ser Val Trp Ser Ser Glu Gln Asp Leu Ala Asn Asp Thr
625 630 635 640

Val Gln Leu Leu Val Thr Leu Val Glu Arg Arg Glu Arg Ala Asn Leu
645 650 655

Val Ile Gln Cys Glu Asn Trp Trp Asn Leu Ala Lys Gln Phe Ala Ser
660 665 670

Arg Ser Pro Pro Leu Asn Phe Leu Ser Ser Pro Val Gln Arg Thr Leu
675 680 685

Met Lys Ala Leu Val Leu Gly Gly Phe Ala His Met Asp Thr Glu Thr
690 695 700

Lys Gln Gln Tyr Trp Thr Glu Val Leu Gln Pro Leu Gln Gln Arg Phe
705 710 715 720

Leu Arg Val Ile Asn Gln Glu Asn Phe Gln Gln Met Cys Gln Gln Glu
725 730 735

Glu Val Lys Gln Glu Ile Thr Ala Thr Leu Glu Ala Leu Cys Gly Ile
740 745 750

Ala Glu Ala Thr Gln Ile Asp Asn Val Ala Ile Leu Phe Asn Phe Leu
755 760 765

Met Asp Phe Leu Thr Asn Cys Ile Gly Leu Met Glu Val Tyr Lys Asn
770 775 780

Thr Pro Glu Thr Val Asn Leu Ile Ile Glu Val Phe Val Glu Val Ala
785 790 795 800

His Lys Gln Ile Cys Tyr Leu Gly Glu Ser Lys Ala Met Asn Leu Tyr
805 810 815

Glu Ala Cys Leu Thr Leu Leu Gln Val Tyr Ser Lys Asn Asn Leu Gly
820 825 830

Arg Gln Arg Ile Asp Val Thr Ala Glu Glu Glu Gln Tyr Gln Asp Leu
835 840 845

Leu Leu Ile Met Glu Leu Leu Thr Asn Leu Leu Ser Lys Glu Phe Ile
850 855 860

Asp Phe Ser Asp Thr Asp Glu Val Phe Arg Gly His Glu Pro Gly Gln
865 870 875 880

Ala Ala Asn Arg Ser Val Ser Ala Ala Asp Val Val Leu Tyr Gly Val
885 890 895

Asn Leu Ile Leu Pro Leu Met Ser Gln Asp Leu Leu Lys Phe Pro Thr
900 905 910

Leu Cys Asn Gln Tyr Tyr Lys Leu Ile Thr Phe Ile Cys Glu Ile Phe
915 920 925

Pro Glu Lys Ile Pro Gln Leu Pro Glu Asp Leu Phe Lys Ser Leu Met
930 935 940

```
Tyr Ser Leu Glu Leu Gly Met Thr Ser Met Ser Ser Glu Val Cys Gln
945                 950                 955                 960

Leu Cys Leu Glu Ala Leu Thr Pro Leu Ala Glu Gln Cys Ala Lys Ala
                965                 970                 975

Gln Glu Thr Asp Ser Pro Leu Phe Leu Ala Thr Arg His Phe Leu Lys
            980                 985                 990

Leu Val Phe Asp Met Leu Val Leu  Gln Lys His Asn Thr  Glu Met Thr
        995                 1000                1005

Thr Ala  Ala Gly Glu Ala Phe  Tyr Thr Leu Val Cys  Leu His Gln
    1010                1015                1020

Ala Glu  Tyr Ser Glu Leu Val  Glu Thr Leu Leu Ser  Ser Gln Gln
    1025                1030                1035

Asp Pro  Val Ile Tyr Gln Arg  Leu Ala Asp Ala Phe  Asn Lys Leu
    1040                1045                1050

Thr Ala  Ser Ser Thr Pro Pro  Thr Leu Asp Arg Lys  Gln Lys Met
    1055                1060                1065

Ala Phe  Leu Lys Ser Leu Glu  Glu Phe Met Ala Asn  Val Gly Gly
    1070                1075                1080

Leu Leu  Cys Val Lys
    1085

<210>  256
<211>  78
<212>  PRT
<213>  Homo sapien

<400>  256

Met Val Leu Met Thr Ser Ser Gly Gln Pro Ser Cys Pro Gly Ile Met
1               5                   10                  15

Ala Cys Gln His Ser Leu Cys Pro Pro Asn Leu Arg Pro Arg Met Arg
            20                  25                  30

Ser Cys Gln His Asn Ile His Pro Phe Glu Gln Met Glu Ser Gly Thr
        35                  40                  45
```

Leu Thr Gln Pro Ser Val Leu Asn Asn Thr Ala Ile Ile Ala Thr Trp
50 55 60

Leu Ser Arg Gln Cys Lys Pro Ser Glu Ser Ala Glu Leu Phe
65 70 75

<210> 257
<211> 595
<212> PRT
<213> Homo sapien

<400> 257

Val Gln Lys Thr Asn Gln Cys Leu Gln Gly Gln Ser Leu Lys Thr Ser
1 5 10 15

Leu Thr Leu Lys Val Asp Arg Gly Ser Glu Glu Thr Tyr Arg Pro Glu
20 25 30

Phe Pro Ser Thr Lys Gly Leu Val Arg Ser Leu Ala Glu Gln Phe Gln
35 40 45

Arg Met Gln Gly Val Ser Met Arg Asp Ser Thr Gly Phe Lys Asp Arg
50 55 60

Ser Leu Ser Gly Ser Leu Arg Lys Asn Ser Ser Pro Ser Asp Ser Lys
65 70 75 80

Pro Pro Phe Ser Gln Gly Gln Glu Lys Gly His Trp Pro Trp Ala Lys
85 90 95

Gln Gln Ser Ser Leu Glu Gly Gly Asp Arg Pro Leu Ser Trp Glu Glu
100 105 110

Ser Thr Glu His Ser Ser Leu Ala Leu Asn Ser Gly Leu Pro Asn Gly
115 120 125

Glu Thr Ser Ser Gly Gly Gln Pro Arg Leu Ala Glu Pro Asp Ile Tyr
130 135 140

Gln Glu Lys Leu Ser Gln Val Arg Asp Val Arg Ser Lys Asp Leu Gly
145 150 155 160

Ser Ser Thr Asp Leu Gly Thr Ser Leu Pro Leu Asp Ser Trp Val Asn
165 170 175

Ile Thr Arg Phe Cys Asp Ser Gln Leu Lys His Gly Ala Pro Arg Pro
180 185 190

Gly Met Lys Ser Ser Pro His Asp Ser His Thr Cys Val Thr Tyr Pro
195 200 205

Glu Arg Asn His Ile Leu Leu His Pro His Trp Asn Gln Asp Thr Glu
210 215 220

Gln Glu Thr Ser Glu Leu Glu Ser Leu Tyr Gln Ala Ser Leu Gln Ala
225 230 235 240

Ser Gln Ala Gly Cys Ser Gly Trp Gly Gln Gln Asp Thr Ala Trp His
245 250 255

Pro Leu Ser Gln Thr Gly Ser Ala Asp Gly Met Gly Arg Arg Leu His
260 265 270

Ser Ala His Asp Pro Gly Leu Ser Lys Thr Ser Thr Ala Glu Met Glu
275 280 285

His Gly Leu His Glu Ala Arg Thr Val Arg Thr Ser Gln Ala Thr Pro
290 295 300

Cys Arg Gly Leu Ser Arg Glu Cys Gly Glu Asp Glu Gln Tyr Ser Ala
305 310 315 320

Glu Asn Leu Arg Arg Ile Ser Arg Ser Leu Ser Gly Thr Val Val Ser
325 330 335

Glu Arg Glu Glu Ala Pro Val Ser Ser His Ser Phe Asp Ser Ser Asn
340 345 350

Val Arg Lys Pro Leu Glu Thr Gly His Arg Cys Ser Ser Ser Ser Ser
355 360 365

Leu Pro Val Ile His Asp Pro Ser Val Phe Leu Leu Gly Pro Gln Leu
370 375 380

Tyr Leu Pro Gln Pro Gln Phe Leu Ser Pro Asp Val Leu Met Pro Thr
385 390 395 400

Met Ala Gly Glu Pro Asn Arg Leu Pro Gly Thr Ser Arg Ser Val Gln
405 410 415

Gln Phe Leu Ala Met Cys Asp Arg Gly Glu Thr Ser Gln Gly Ala Lys
420 425 430

Tyr Thr Gly Arg Thr Leu Asn Tyr Gln Ser Leu Pro His Arg Ser Arg
435 440 445

Thr Asp Asn Ser Trp Ala Pro Trp Ser Glu Thr Asn Gln His Ile Gly
450 455 460

Thr Arg Phe Leu Thr Thr Pro Gly Cys Asn Pro Gln Leu Thr Tyr Thr
465 470 475 480

Ala Thr Leu Pro Glu Arg Ser Lys Gly Leu Gln Val Pro His Thr Gln
485 490 495

Ser Trp Ser Asp Leu Phe His Ser Pro Ser His Pro Pro Ile Val His
500 505 510

Pro Val Tyr Pro Pro Ser Ser Ser Leu His Val Pro Leu Arg Ser Ala
515 520 525

Trp Asn Ser Asp Pro Val Pro Gly Ser Arg Thr Pro Gly Pro Arg Arg
530 535 540

Val Asp Met Pro Pro Asp Asp Asp Trp Arg Gln Ser Ser Tyr Ala Ser
545 550 555 560

His Ser Gly His Arg Arg Thr Val Gly Glu Gly Phe Leu Phe Val Leu
565 570 575

Ser Asp Ala Pro Arg Arg Glu Gln Ile Arg Ala Arg Val Leu Gln His
580 585 590

Ser Gln Trp
595

<210> 258
<211> 55
<212> PRT
<213> Homo sapien

<400> 258

Met Thr Val Met Ile Leu Leu Phe Lys Lys Asn Pro Asn Cys Tyr Phe
1 5 10 15

Asp Leu Tyr Asp Leu Thr Leu Asn His Gly Ser Ile Thr Met Met Phe
20 25 30

Lys Thr Leu Ile Asp Ser Thr Cys Phe Lys Asn Ser Gln Ile Pro Ser
35 40 45

Ala Phe Ile Ile Arg Asp Arg
50 55

<210> 259
<211> 43
<212> PRT
<213> Homo sapien

<400> 259

Met Met Leu Thr Met Glu Phe Lys Asn Lys Gln Gln His Phe Val Val
1 5 10 15

Ser Thr Gly Val Gly Val Glu Glu Leu Gln Arg His His Gly Asn Lys
20 25 30

Ser Leu Pro Arg Ile Ser Gly Pro Arg Asn Leu
35 40

<210> 260
<211> 75
<212> PRT
<213> Homo sapien

<400> 260

Met Ala Tyr Arg Met Lys Arg Gly Thr Arg Asn Pro Cys Gly Arg Gly
1 5 10 15

Leu Asp Leu Lys Gln Cys Pro Leu Trp Leu Leu Leu Pro Trp Leu Thr
20 25 30

Gly Phe Leu Asp His Val His Phe Thr Gly Pro Trp Asp Leu His Leu
35 40 45

Leu Ala Ser Pro Ala Gly Leu Ile Pro Ala Arg Ala Pro Ser Phe Leu
50 55 60

Leu Met Val Phe Arg Trp Pro Asp His Gly Lys
65 70 75

<210> 261
<211> 218
<212> PRT
<213> Homo sapien

<400> 261

Met Ile Asn His Leu Ser Pro His Gln Ala Ala Ala Pro Val Asp Gln
1 5 10 15

Thr Pro Arg Thr Leu Ala Thr Met Gly Gln Arg Ala Leu Pro Ser Ser
20 25 30

Leu Ala Leu Leu Ser Arg Pro Leu Ser Pro Pro Pro Ala Ala Cys Ser
35 40 45

Gly Asp Pro Gly Cys Gly Ser Gly Ala Gly Leu Pro Ser Ala Ser Ala
50 55 60

Ala Ala Gly Ile Ala Ser Ser Ala Val Glu Ala Val Cys Gly Asp Ala
65 70 75 80

Ala Pro Ala Cys Leu Leu Arg Thr Pro Leu Arg Gly Leu Leu Lys Pro
85 90 95

Thr Gly Pro Arg Ser Thr Met Glu Cys Pro Pro Ala Leu Ile Val Gln
100 105 110

Pro Pro Ala Gly Gly Met Ala Arg Arg Ala Ala Ser Gln Pro Trp Ala
115 120 125

Ala Ala Ser Ala Thr Pro Met Leu Ser Ser Lys Ala Ser Leu Cys Ile
130 135 140

Pro Thr Glu Arg Pro Pro Pro Gln Pro Leu Met Arg Thr Pro Ala Ala
145 150 155 160

Arg Ser His Trp Pro Ile Pro His Pro Ala Ser Thr Ala Cys Pro Ala
165 170 175

Pro Leu Pro Val Val Leu Val Ala Pro Arg Ser Thr Ile Leu Ser Met
180 185 190

Ser Arg Thr Trp Thr Cys Arg Arg Trp Ala Val Ala Pro Cys Arg Ala
195 200 205

```
Glu Lys Leu Met Cys Ser Ser Ser Arg Ser
    210                 215


<210>  262
<211>  104
<212>  PRT
<213>  Homo sapien

<400>  262

Met Pro Ser Phe Phe Cys Phe Ser Ile Ser Leu Ile Arg Asp Trp Lys
1               5                   10                  15


Val Ser Ile Arg Ser Asn Thr Asp Phe Ile Val Ile Gly Thr Asn Cys
            20                  25                  30


Ser Pro Thr Thr Pro Tyr Ser Ala Ser Ser Ile Thr Leu Leu Cys Glu
        35                  40                  45


Ile Leu Arg Asn Gly Leu Pro Leu Gln Gly Leu Asn Leu Pro Tyr Leu
    50                  55                  60


Arg Phe Glu Ser Ser Val Leu Phe Cys Ile Cys Phe Lys Tyr Leu Gly
65                  70                  75                  80


Ser Val Thr His Ala Asn Met Thr Cys Pro Val Gln Ala Thr Leu Gly
                85                  90                  95


Ile His Ile Ser His Val Ser Ser
            100


<210>  263
<211>  260
<212>  PRT
<213>  Homo sapien

<400>  263

Glu Lys Lys Lys Lys Met Lys Asn Glu Asn Ala Asp Lys Leu Leu Lys
1               5                   10                  15


Ser Glu Lys Gln Met Lys Lys Ser Glu Lys Lys Ser Lys Gln Glu Lys
            20                  25                  30


Glu Lys Ser Lys Lys Lys Lys Gly Gly Lys Thr Glu Gln Asp Gly Tyr
        35                  40                  45
```

```
Gln Lys Pro Thr Asn Lys His Phe Thr Gln Ser Pro Lys Lys Ser Val
    50                  55                  60

Ala Asp Leu Leu Gly Ser Phe Glu Gly Lys Arg Arg Leu Leu Leu Ile
65                  70                  75                  80

Thr Ala Pro Lys Ala Glu Asn Asn Met Tyr Val Gln Gln Arg Asp Glu
                85                  90                  95

Tyr Leu Glu Ser Phe Cys Lys Met Ala Thr Arg Lys Ile Ser Val Ile
            100                 105                 110

Thr Ile Phe Gly Pro Val Asn Asn Ser Thr Met Lys Ile Asp His Phe
        115                 120                 125

Gln Leu Asp Asn Glu Lys Pro Met Arg Val Val Asp Asp Glu Asp Leu
    130                 135                 140

Val Asp Gln Arg Leu Ile Ser Glu Leu Arg Lys Glu Tyr Gly Met Thr
145                 150                 155                 160

Tyr Asn Asp Phe Phe Met Val Leu Thr Asp Val Asp Leu Arg Val Lys
                165                 170                 175

Gln Tyr Tyr Glu Val Pro Ile Thr Met Lys Ser Val Phe Asp Leu Ile
            180                 185                 190

Asp Thr Phe Gln Ser Arg Ile Lys Asp Met Glu Lys Gln Lys Lys Glu
        195                 200                 205

Gly Ile Val Cys Lys Glu Asp Lys Lys Gln Ser Leu Glu Asn Phe Leu
    210                 215                 220

Ser Arg Phe Arg Trp Arg Arg Arg Leu Leu Val Ile Ser Ala Pro Asn
225                 230                 235                 240

Asp Glu Asp Trp Ala Tyr Ser Gln Gln Leu Ser Ala Leu Ser Gly Gln
                245                 250                 255

Ala Cys Thr Leu
            260
```

<210> 264
<211> 62
<212> PRT

<213> Homo sapien

<400> 264

Met Ser Gly Phe Ile Tyr Val Leu Glu Lys Asp His Leu Lys Lys Ile
1 5 10 15

Asn Thr Phe Ser Thr Thr Lys Lys Lys Lys Lys Lys Lys Lys Lys Lys
20 25 30

Arg Arg Gly Gly Glu Pro Gly Ala Gln Ser Gly Pro Arg Gly Ala Asn
35 40 45

Trp Val Leu Pro Ala His Ile Pro Pro Lys Tyr Trp His Thr
50 55 60

<210> 265
<211> 89
<212> PRT
<213> Homo sapien

<400> 265

Met Leu Gln Leu Asn Thr Arg Phe Tyr Phe Leu Ser Asn Cys Gly Phe
1 5 10 15

Val Phe Ile Tyr His Pro Leu Phe Ile Pro Phe Leu Thr His Thr Leu
20 25 30

Cys Arg Ala Ser Gly Ile Tyr Tyr Ser Thr Val Cys Leu Cys Lys Arg
35 40 45

Leu Ser Val Leu Ala Ser Thr Tyr Glu Arg Met His Ala Lys Phe Cys
50 55 60

Leu Ser Met Pro Gly Leu Ile Ser Leu Lys Gln Asn Asp Leu Arg Val
65 70 75 80

Pro Ser Met Leu Phe Ile Leu Pro Asn
85

<210> 266
<211> 38
<212> PRT
<213> Homo sapien

<400> 266

Met Thr Ser Arg Trp Leu Asn Phe Ser Cys Leu Trp Cys Phe Gly Pro

1 5 10 15

Asn Ser Thr Gly Gln His His Asp His Met Glu Thr Tyr Phe Trp Lys
20 25 30

Gln Asn Phe Asn Phe Ile
35

<210> 267
<211> 111
<212> PRT
<213> Homo sapien

<400> 267

Asn Asp Leu Asp Arg Tyr Asn Pro Leu Ser Ser Gln Arg Leu Val Arg
1 5 10 15

Asn Ala Leu Ala His Val Gly Ala Lys Glu Arg Glu Leu Ser Trp Ala
20 25 30

His Ser Glu Ser Phe Ala Ala Leu Cys Arg Tyr Gly Lys Arg Glu Phe
35 40 45

Lys Ile Gly Gly Glu Leu Arg Ile Gly Lys Gln Pro Tyr Arg Leu Gln
50 55 60

Ile Gln Leu Ser Ala Gln Arg Ser His Thr Leu Glu Phe Gln Ser Leu
65 70 75 80

Glu Asp Leu Ile Met Gly Glu Ala Thr Gln Arg Pro Arg Ser Gly Ala
85 90 95

Arg Pro Val Leu Gln Glu Leu Ala Thr His Leu His Pro Ala Glu
100 105 110

<210> 268
<211> 60
<212> PRT
<213> Homo sapien

<400> 268

Met Val Asn Thr Val Leu Leu Ser Leu Lys Ile Ser Leu Phe Cys Pro
1 5 10 15

His Gln Leu Phe Tyr Cys Ser Val Leu Arg Lys Pro Asn Ser Cys Val
20 25 30

Phe Phe Pro Ser Leu Leu Ile Leu Ser Cys Val Pro Ser Gly Lys Cys
35 40 45

His Tyr Phe Leu Asp Ile Leu Asn Leu Leu Phe Leu
50 55 60

<210> 269
<211> 72
<212> PRT
<213> Homo sapien

<400> 269

Met Cys Leu Cys Ile Leu Val Ser Lys Leu Arg Thr Ser Asp Glu Leu
1 5 10 15

Pro Val Val Pro Ser Tyr Cys Arg Arg Leu Glu Val Arg Gly Ile Ser
20 25 30

Ala Ser Thr Arg Glu Ala Glu Val Ala Ser Glu Pro Thr Ile Met Thr
35 40 45

Ala Cys Thr Pro Ser Leu Ala Thr Val Arg Glu Leu Leu Ser Gln Ile
50 55 60

Lys Arg Lys Gln Ser Leu Leu Ser
65 70

<210> 270
<211> 152
<212> PRT
<213> Homo sapien

<400> 270

Gly Ser Leu Gly Gly Glu Pro Gly Val Ser Cys Leu Lys Met His Ser
1 5 10 15

Asp Ala Ala Ala Val Asn Phe Gln Leu Asn Ser His Leu Ser Thr Leu
20 25 30

Ala Asn Ile His Lys Ile Tyr His Thr Leu Asn Lys Leu Asn Leu Thr
35 40 45

Glu Asp Ile Gly Gln Asp Asp His Gln Thr Gly Ser Leu Arg Ser Cys
50 55 60

Ser Ser Ser Asp Cys Phe Asn Lys Val Met Pro Pro Arg Lys Lys Arg
65 70 75 80

Arg Pro Ala Ser Gly Asp Asp Leu Ser Ala Lys Lys Ser Arg His Asp
85 90 95

Ser Met Tyr Arg Lys Tyr Asp Ser Thr Arg Ile Lys Thr Glu Glu Glu
100 105 110

Ala Phe Ser Ser Lys Arg Cys Leu Glu Trp Phe Tyr Glu Tyr Ala Gly
115 120 125

Thr Asp Asp Val Val Gly Pro Glu Gly Met Glu Lys Phe Cys Glu Asp
130 135 140

Ile Gly Val Glu Pro Glu Asn Val
145 150

<210> 271
<211> 52
<212> PRT
<213> Homo sapien

<400> 271

Met Glu Pro His Ile Met Lys Phe Asn Ser His Val Lys Thr Phe Cys
1 5 10 15

Ile Val Gly Cys Gln Lys Tyr Leu Pro Lys Leu Ser Phe Asp Leu Ser
20 25 30

Glu Trp Gly Trp Leu Leu Pro Ile Leu Gln Phe Val Ser Gln Ala Trp
35 40 45

Arg Asn Gln Ala
50

<210> 272
<211> 449
<212> PRT
<213> Homo sapien

<400> 272

Met Val Met Glu Lys Pro Ser Pro Leu Leu Val Gly Arg Glu Phe Val
1 5 10 15

Arg Gln Tyr Tyr Thr Leu Leu Asn Lys Ala Pro Glu Tyr Leu His Arg
20 25 30

Phe Tyr Gly Arg Asn Ser Ser Tyr Val His Gly Gly Val Asp Ala Ser
35 40 45

Gly Lys Pro Gln Glu Ala Val Tyr Gly Gln Asn Asp Ile His His Lys
50 55 60

Val Leu Ser Leu Asn Phe Ser Glu Cys His Thr Lys Ile Arg His Val
65 70 75 80

Asp Ala His Ala Thr Leu Ser Asp Gly Val Val Val Gln Val Met Gly
85 90 95

Leu Leu Ser Asn Ser Gly Gln Pro Glu Arg Lys Phe Met Gln Thr Phe
100 105 110

Val Leu Ala Pro Glu Gly Ser Val Pro Asn Lys Phe Tyr Val His Asn
115 120 125

Asp Met Phe Arg Tyr Glu Asp Glu Val Phe Gly Asp Ser Glu Pro Glu
130 135 140

Leu Asp Glu Glu Ser Glu Asp Glu Val Glu Glu Glu Gln Glu Glu Arg
145 150 155 160

Gln Pro Ser Pro Glu Pro Val Gln Glu Asn Ala Asn Ser Gly Tyr Tyr
165 170 175

Glu Ala His Pro Val Thr Asn Gly Ile Glu Glu Pro Leu Glu Glu Ser
180 185 190

Ser His Glu Pro Glu Pro Glu Pro Glu Ser Glu Thr Lys Thr Glu Glu
195 200 205

Leu Lys Pro Gln Val Glu Glu Lys Asn Leu Glu Glu Leu Glu Glu Lys
210 215 220

Ser Thr Thr Pro Pro Pro Ala Glu Pro Val Ser Leu Pro Gln Glu Pro
225 230 235 240

Pro Lys Pro Arg Val Glu Ala Lys Pro Glu Val Gln Ser Gln Pro Pro
245 250 255

Arg Val Arg Glu Gln Arg Pro Arg Glu Arg Pro Gly Phe Pro Pro Arg
260 265 270

Gly Pro Arg Pro Gly Arg Gly Asp Met Glu Gln Asn Asp Ser Asp Asn
275 280 285

Arg Arg Ile Ile Arg Tyr Pro Asp Ser His Gln Leu Phe Val Gly Asn
290 295 300

Leu Pro His Asp Ile Asp Glu Asn Glu Leu Lys Glu Phe Phe Met Ser
305 310 315 320

Phe Gly Asn Val Val Glu Leu Arg Ile Asn Thr Lys Gly Val Gly Gly
325 330 335

Lys Leu Pro Asn Phe Gly Phe Val Val Phe Asp Asp Ser Glu Pro Val
340 345 350

Gln Arg Ile Leu Ile Ala Lys Pro Ile Met Phe Arg Gly Glu Val Arg
355 360 365

Leu Asn Val Glu Glu Lys Lys Thr Arg Ala Ala Arg Glu Arg Glu Thr
370 375 380

Arg Gly Gly Gly Asp Asp Arg Arg Asp Ile Arg Arg Asn Asp Arg Gly
385 390 395 400

Pro Gly Gly Pro Arg Gly Ile Val Gly Gly Gly Met Met Arg Asp Arg
405 410 415

Asp Gly Arg Gly Pro Pro Pro Arg Gly Gly Met Ala Gln Lys Leu Gly
420 425 430

Ser Gly Arg Gly Thr Gly Gln Met Glu Gly Arg Phe Thr Gly Gln Arg
435 440 445

Arg

<210> 273
<211> 63
<212> PRT
<213> Homo sapien

<400> 273

Met Cys Cys Asp Val Ser Glu Arg Ala Glu Phe Arg Leu Val Ser Ala
1 5 10 15

Arg Cys Ser Phe Ser His Pro Arg Thr Val Ala Arg Leu Leu Leu Arg
20 25 30

His Pro Gly Gln Leu Pro Leu Pro Phe Gln Trp Gly Leu Thr Trp Leu
35 40 45

Pro Ser Leu Ala Ala Asn Arg Arg Ala Pro Gln His Ser Arg Ser
50 55 60

<210> 274
<211> 60
<212> PRT
<213> Homo sapien

<400> 274

Met Asp Pro Gly Arg Tyr Cys Leu Val Leu Gln Glu Leu Met Gln Phe
1 5 10 15

His Ser Glu Ala Cys Lys Ile Leu Asn Phe Arg Asp Asn Arg Pro Asp
20 25 30

Thr Phe Leu Ile Ser Phe Tyr Ser Leu Met Ser Asn Asn Thr Ile Phe
35 40 45

Lys Asn Met Val Leu Ile Cys Leu Ala Ser Asn Leu
50 55 60

<210> 275
<211> 111
<212> PRT
<213> Homo sapien

<400> 275

Lys Leu Ile Val Tyr Pro Pro Pro Pro Ala Lys Gly Gly Ile Ser Val
1 5 10 15

Thr Asn Glu Asp Leu His Cys Leu Asn Glu Gly Glu Phe Leu Asn Asp
20 25 30

Val Ile Ile Asp Phe Tyr Leu Lys Tyr Leu Val Leu Glu Lys Leu Lys
35 40 45

Lys Glu Asp Ala Asp Arg Ile His Ile Phe Ser Ser Phe Phe Tyr Lys
    50   55   60

Arg Leu Asn Gln Arg Glu Arg Arg Asn His Glu Thr Thr Asn Leu Ser
65   70   75   80

Ile Gln Gln Lys Arg His Gly Arg Val Lys Thr Trp Thr Arg His Val
                85   90   95

Asp Ile Phe Glu Lys Asp Phe Ile Phe Val Pro Leu Asn Glu Ala
            100   105   110

<210> 276
<211> 97
<212> PRT
<213> Homo sapien

<400> 276

Met Ser Gln Asp Thr Ser Arg Ser Gln Glu Arg Ala Ala Gly Pro Gln
1   5   10   15

Arg Thr Arg Arg Arg Pro Arg Thr Trp Ser Gly Gly Val Glu Pro Thr
            20   25   30

Ala Ala Ala Pro Trp Ala Ala Ala Met Ala His Thr Gly Arg His Gly
        35   40   45

Ser Gly Ala Ala Ala Thr Ala Ser Ser Thr Arg Gly Asp Gly Ala Ala
    50   55   60

Arg Arg Gly Ala Ala Arg Gly Thr Asp Ala Ala Glu Arg Arg Arg Ala
65   70   75   80

Ala Ser Arg Gly Ala Ala Glu Pro Lys Ala Thr Ala Ser Gly Gly Gly
                85   90   95

Gly

<210> 277
<211> 76
<212> PRT
<213> Homo sapien

<400> 277

Met Gly Ser Cys Pro Leu Trp Val Arg Ser Ser Thr Cys Arg Val Glu

1 5 10 15

Val Gly Tyr Val His Thr Phe Asn Asp Asn Leu His Ile Ser Ala Pro
20 25 30

Thr Gly Pro Lys Leu Phe Leu Gly Phe Lys Val Val Val Cys Leu Phe
35 40 45

Phe Ser Phe Phe Phe Phe Phe Phe Phe Phe Gly Glu Val Glu Phe Gly
50 55 60

Ser Gly Trp Pro Arg Cys Gly Val Cys Lys Gly Arg
65 70 75

<210> 278
<211> 20
<212> PRT
<213> Homo sapien

<400> 278

Met Glu Asp Gln Ile Ile Leu Asn Tyr Ile Ser Ile Val Pro Gly Lys
1 5 10 15

Thr Gln Val Leu
20

<210> 279
<211> 24
<212> PRT
<213> Homo sapien

<400> 279

Met Val His Leu Met His Ala Arg Ala Arg Ala Ser Cys Asp Gly Cys
1 5 10 15

Val Val Ala Ala Glu Val His Val
20

<210> 280
<211> 101
<212> PRT
<213> Homo sapien

<400> 280

Leu Phe Phe Phe Lys Lys Phe Ile Leu Arg Trp Ser Leu Thr Leu Ser
1 5 10 15

Leu Arg Leu Glu Cys Ser Asp Ser Ile Ser Ala His Cys Asn Leu Arg
20 25 30

Leu Pro Gly Leu Ser Asn Phe Cys Ala Ser Ala Ser Gln Val Ser Glu
35 40 45

Ile Thr Gly Val Cys His His Thr Gln Leu Phe Phe Ile Phe Tyr Phe
50 55 60

Ala Ala Lys Met Gly Phe Arg His Val Gly Arg Thr Gly Leu Glu Leu
65 70 75 80

Leu Ala Ser Ser Gly Pro Pro Thr Ser Ala Ser Gln Ser Ala Gly Ile
85 90 95

Thr Gly Val Ser His
100

<210> 281
<211> 43
<212> PRT
<213> Homo sapien

<400> 281

Met Trp Gly His Gly Leu Asp Asp Gly Leu His Arg Ser Phe His Leu
1 5 10 15

Cys Glu Ser Lys Ser Gly Gln Ser Ala Arg Thr Gln Ser Leu Thr Leu
20 25 30

Gly Gln Leu Leu Arg Thr Asn Pro Gln His Leu
35 40

<210> 282
<211> 46
<212> PRT
<213> Homo sapien

<400> 282

Met Ala Gly Asn Ile His Pro Gly Thr Phe Gly Pro Gly Ser Pro His
1 5 10 15

Leu Phe Phe Leu Cys Gly Val Val Ala Phe Phe Leu Phe Ile Val Ala
20 25 30

Arg Glu Ala Lys Ile Tyr Ser Phe Ser Met Asn Pro Asn Met
35 40 45

<210> 283
<211> 70
<212> PRT
<213> Homo sapien

<400> 283

Met Pro Gly Ser His Leu Cys Met Phe Asn Thr Val Thr His Asp Val
1 5 10 15

Ile Thr Glu Trp Arg Arg Trp Lys Gly Pro Cys Arg Ser Phe Ser Trp
20 25 30

His Pro Asn Phe Thr Glu Gly Glu Leu Arg Pro Glu Leu Arg Asp Val
35 40 45

Leu Arg Ile Pro Glu Ser His Ser Ser Val Arg Ser Val Ile His Lys
50 55 60

Glu Val Ile Ile Lys Val
65 70

<210> 284
<211> 49
<212> PRT
<213> Homo sapien

<400> 284

Met Ser Ser Ser Leu Phe Ala Phe Leu Leu Thr Tyr Phe Val Val Phe
1 5 10 15

Lys Asp Cys Ala Gly Asp Ile Leu Glu Gly Ile Asn Gly Leu His Ser
20 25 30

Lys Arg Cys Gly Leu Ser Lys Leu Phe Ser Val Phe Ile Thr Glu Thr
35 40 45

Asp

<210> 285
<211> 1544
<212> PRT
<213> Homo sapien

<400> 285

Met Tyr Ala Ala Val Glu His Gly Pro Val Leu Cys Ser Asp Ser Asn
1 5 10 15

Ile Leu Cys Leu Ser Trp Lys Gly Arg Val Pro Lys Ser Glu Lys Glu
20 25 30

Lys Pro Val Cys Arg Arg Arg Tyr Tyr Glu Glu Gly Trp Leu Ala Thr
35 40 45

Gly Asn Gly Arg Gly Val Val Gly Val Thr Phe Thr Ser Ser His Cys
50 55 60

Arg Arg Asp Arg Ser Thr Pro Gln Arg Ile Asn Phe Asn Leu Arg Gly
65 70 75 80

His Asn Ser Glu Val Val Leu Val Arg Trp Asn Glu Pro Tyr Gln Lys
85 90 95

Leu Ala Thr Cys Asp Ala Asp Gly Gly Ile Phe Val Trp Ile Gln Tyr
100 105 110

Glu Gly Arg Trp Ser Val Glu Leu Val Asn Asp Arg Gly Ala Gln Val
115 120 125

Ser Asp Phe Thr Trp Ser His Asp Gly Thr Gln Ala Leu Ile Ser Tyr
130 135 140

Arg Asp Gly Phe Val Leu Val Gly Ser Val Ser Gly Gln Arg His Trp
145 150 155 160

Ser Ser Glu Ile Asn Leu Glu Ser Gln Ile Thr Cys Gly Ile Trp Thr
165 170 175

Pro Asp Asp Gln Gln Val Leu Phe Gly Thr Ala Asp Gly Gln Val Ile
180 185 190

Val Met Asp Cys His Gly Arg Met Leu Ala His Val Leu Leu His Glu
195 200 205

Ser Asp Gly Val Leu Gly Met Ser Trp Asn Tyr Pro Ile Phe Leu Val
210 215 220

Glu Asp Ser Ser Glu Ser Asp Thr Asp Ser Asp Asp Tyr Ala Pro Pro

225 230 235 240

Gln Asp Gly Pro Ala Ala Tyr Pro Ile Pro Val Gln Asn Ile Lys Pro
245 250 255

Leu Leu Thr Val Ser Phe Thr Ser Gly Asp Ile Ser Leu Met Asn Asn
260 265 270

Tyr Asp Asp Leu Ser Pro Thr Val Ile Arg Ser Gly Leu Lys Glu Val
275 280 285

Val Ala Gln Trp Cys Thr Gln Gly Asp Leu Leu Ala Val Ala Gly Met
290 295 300

Glu Arg Gln Thr Gln Leu Gly Glu Leu Pro Asn Gly Pro Leu Leu Lys
305 310 315 320

Ser Ala Met Val Lys Phe Tyr Asn Val Arg Gly Glu His Ile Phe Thr
325 330 335

Leu Asp Thr Leu Val Gln Arg Pro Ile Ile Ser Ile Cys Trp Gly His
340 345 350

Arg Asp Ser Arg Leu Leu Met Ala Ser Gly Pro Ala Leu Tyr Val Val
355 360 365

Arg Val Glu His Arg Val Ser Ser Leu Gln Leu Leu Cys Gln Gln Ala
370 375 380

Ile Ala Ser Thr Leu Arg Glu Asp Lys Asp Val Ser Lys Leu Thr Leu
385 390 395 400

Pro Pro Arg Leu Cys Ser Tyr Leu Ser Thr Ala Phe Ile Pro Thr Ile
405 410 415

Lys Pro Pro Ile Pro Asp Pro Asn Asn Met Arg Asp Phe Val Ser Tyr
420 425 430

Pro Ser Ala Gly Asn Glu Arg Leu His Cys Thr Met Lys Arg Thr Glu
435 440 445

Asp Asp Pro Glu Val Gly Gly Pro Cys Tyr Thr Leu Tyr Leu Glu Tyr
450 455 460

```
Leu Gly Gly Leu Val Pro Ile Leu Lys Gly Arg Arg Ile Ser Lys Leu
465                 470                 475                 480

Arg Pro Glu Phe Val Ile Met Asp Pro Arg Thr Asp Ser Lys Pro Asp
                485                 490                 495

Glu Ile Tyr Gly Asn Ser Leu Ile Ser Thr Val Ile Asp Ser Cys Asn
            500                 505                 510

Cys Ser Asp Ser Ser Asp Ile Glu Leu Ser Asp Asp Trp Ala Ala Lys
        515                 520                 525

Lys Ser Pro Lys Ile Ser Arg Ala Ser Lys Ser Pro Lys Leu Pro Arg
    530                 535                 540

Ile Ser Ile Glu Ala Arg Lys Ser Pro Lys Leu Pro Arg Ala Ala Gln
545                 550                 555                 560

Glu Leu Ser Arg Ser Pro Arg Leu Pro Leu Arg Lys Pro Ser Val Gly
                565                 570                 575

Ser Pro Ser Leu Thr Arg Arg Glu Phe Pro Phe Glu Asp Ile Thr Gln
            580                 585                 590

His Asn Tyr Leu Ala Gln Val Thr Ser Asn Ile Trp Gly Thr Lys Phe
        595                 600                 605

Lys Ile Val Gly Leu Ala Ala Phe Leu Pro Thr Asn Leu Gly Ala Val
    610                 615                 620

Ile Tyr Lys Thr Ser Leu Leu His Leu Gln Pro Arg Gln Met Thr Ile
625                 630                 635                 640

Tyr Leu Pro Glu Val Arg Lys Ile Ser Met Asp Tyr Ile Asn Leu Pro
                645                 650                 655

Val Phe Asn Pro Asn Val Phe Ser Glu Asp Glu Asp Asp Leu Pro Val
            660                 665                 670

Thr Gly Ala Ser Gly Val Pro Glu Asn Ser Pro Pro Cys Thr Val Asn
        675                 680                 685

Ile Pro Ile Ala Pro Ile His Ser Ser Ala Gln Ala Met Ser Pro Thr
    690                 695                 700
```

Gln Ser Ile Gly Leu Val Gln Ser Leu Leu Ala Asn Gln Asn Val Gln
705 710 715 720

Leu Asp Val Leu Thr Asn Gln Thr Thr Ala Val Gly Thr Ala Glu His
725 730 735

Ala Gly Asp Arg Cys His Pro Val Thr Gln Val Ser Asn Arg Tyr Ser
740 745 750

Asn Pro Gly Gln Val Ile Phe Gly Ser Val Glu Met Gly Arg Ile Ile
755 760 765

Gln Asn Pro Pro Pro Leu Ser Leu Pro Pro Pro Pro Gln Gly Pro Met
770 775 780

Gln Leu Ser Thr Val Gly His Gly Asp Arg Asp His Glu His Leu Gln
785 790 795 800

Lys Ser Ala Lys Ala Leu Arg Pro Thr Pro Gln Leu Ala Ala Glu Gly
805 810 815

Asp Ala Val Val Phe Ser Ala Pro Gln Glu Val Gln Val Thr Lys Ile
820 825 830

Asn Pro Pro Pro Pro Tyr Pro Gly Thr Ile Pro Ala Ala Pro Thr Thr
835 840 845

Ala Ala Pro Pro Pro Pro Leu Pro Pro Pro Gln Pro Pro Val Asp Val
850 855 860

Cys Leu Lys Lys Gly Asp Phe Ser Leu Tyr Pro Thr Ser Val His Tyr
865 870 875 880

Gln Thr Pro Leu Gly Tyr Glu Arg Ile Thr Thr Phe Asp Ser Ser Gly
885 890 895

Asn Val Glu Glu Val Cys Arg Pro Arg Thr Arg Met Leu Cys Ser Gln
900 905 910

Asn Thr Tyr Thr Leu Pro Gly Pro Gly Ser Ser Ala Thr Leu Arg Leu
915 920 925

Thr Ala Thr Glu Lys Lys Val Pro Gln Pro Cys Ser Ser Ala Thr Leu
930 935 940

Asn Arg Leu Thr Val Pro Arg Tyr Ser Ile Pro Thr Gly Asp Pro Pro
945 950 955 960

Pro Tyr Pro Glu Ile Ala Ser Gln Leu Ala Gln Gly Arg Gly Ala Ala
965 970 975

Gln Arg Ser Asp Asn Ser Leu Ile His Ala Thr Leu Arg Arg Asn Asn
980 985 990

Arg Glu Ala Thr Leu Lys Met Ala Gln Leu Ala Asp Ser Pro Arg Ala
995 1000 1005

Pro Leu Gln Pro Leu Ala Lys Ser Lys Gly Gly Pro Gly Gly Val
1010 1015 1020

Val Thr Gln Leu Pro Ala Arg Pro Pro Pro Ala Leu Tyr Thr Cys
1025 1030 1035

Ser Gln Cys Ser Gly Thr Gly Pro Ser Ser Gln Pro Gly Ala Ser
1040 1045 1050

Leu Ala His Thr Ala Ser Ala Ser Pro Leu Ala Ser Gln Ser Ser
1055 1060 1065

Tyr Ser Leu Leu Ser Pro Pro Asp Ser Ala Arg Asp Arg Thr Asp
1070 1075 1080

Tyr Val Asn Ser Ala Phe Thr Glu Asp Glu Ala Leu Ser Gln His
1085 1090 1095

Cys Gln Leu Glu Lys Pro Leu Arg His Pro Pro Leu Pro Glu Ala
1100 1105 1110

Ala Val Thr Leu Lys Arg Pro Pro Pro Tyr Gln Trp Asp Pro Met
1115 1120 1125

Leu Gly Glu Asp Val Trp Val Pro Gln Glu Arg Thr Ala Gln Thr
1130 1135 1140

Ser Gly Pro Asn Pro Leu Lys Leu Ser Ser Leu Met Leu Ser Gln
1145 1150 1155

Gly Gln His Leu Asp Val Ser Arg Leu Pro Phe Ile Ser Pro Lys

1160 1165 1170

Ser Pro Ala Ser Pro Thr Ala Thr Phe Gln Thr Gly Tyr Gly Met
1175 1180 1185

Gly Val Pro Tyr Pro Gly Ser Tyr Asn Asn Pro Pro Leu Pro Gly
1190 1195 1200

Val Gln Ala Pro Cys Ser Pro Lys Asp Ala Leu Ser Pro Thr Gln
1205 1210 1215

Phe Ala Gln Gln Glu Pro Ala Val Val Leu Gln Pro Leu Tyr Pro
1220 1225 1230

Pro Ser Leu Ser Tyr Cys Thr Leu Pro Pro Met Tyr Pro Gly Ser
1235 1240 1245

Ser Thr Cys Ser Ser Leu Gln Leu Pro Pro Val Ala Leu His Pro
1250 1255 1260

Trp Ser Ser Tyr Ser Ala Cys Pro Pro Met Gln Asn Pro Gln Gly
1265 1270 1275

Thr Leu Pro Pro Lys Pro His Leu Val Val Glu Lys Pro Leu Val
1280 1285 1290

Ser Pro Pro Pro Ala Asp Leu Gln Ser His Leu Gly Thr Glu Val
1295 1300 1305

Met Val Glu Thr Ala Asp Asn Phe Gln Glu Val Leu Ser Leu Thr
1310 1315 1320

Glu Ser Pro Val Pro Gln Arg Thr Glu Lys Phe Gly Lys Lys Asn
1325 1330 1335

Arg Lys Arg Leu Asp Ser Arg Ala Glu Glu Gly Ser Val Gln Ala
1340 1345 1350

Ile Thr Glu Gly Lys Val Lys Lys Glu Ala Arg Thr Leu Ser Asp
1355 1360 1365

Phe Asn Ser Leu Ile Ser Ser Pro His Leu Gly Arg Glu Lys Lys
1370 1375 1380

Lys Val Lys Ser Gln Lys Asp Gln Leu Lys Ser Lys Lys Leu Asn
1385 1390 1395

Lys Thr Asn Glu Phe Gln Asp Ser Ser Glu Ser Glu Pro Glu Leu
1400 1405 1410

Phe Ile Ser Gly Asp Glu Leu Met Asn Gln Ser Gln Gly Ser Arg
1415 1420 1425

Lys Gly Trp Lys Ser Lys Arg Ser Pro Arg Ala Ala Gly Glu Leu
1430 1435 1440

Glu Glu Ala Lys Cys Arg Arg Ala Ser Glu Lys Glu Asp Gly Arg
1445 1450 1455

Leu Gly Ser Gln Gly Phe Val Tyr Val Met Ala Asn Lys Gln Pro
1460 1465 1470

Leu Trp Asn Glu Ala Thr Gln Val Tyr Gln Leu Asp Phe Gly Gly
1475 1480 1485

Arg Val Thr Gln Glu Ser Ala Lys Asn Phe Gln Ile Glu Leu Glu
1490 1495 1500

Gly Arg Gln Val Met Gln Phe Gly Arg Ile Asp Gly Ser Ala Tyr
1505 1510 1515

Ile Leu Asp Phe Gln Tyr Pro Phe Ser Ala Val Gln Ala Phe Ala
1520 1525 1530

Val Ala Leu Ala Asn Val Thr Gln Arg Leu Lys
1535 1540

<210> 286
<211> 56
<212> PRT
<213> Homo sapien

<400> 286

Met Gly Asn Gly Ala Thr Gln Lys Gln Leu Pro Asn Leu Arg Asn Asn
1 5 10 15

Ser Phe Val Val Tyr Phe Leu Val Leu Val Gly Ala Leu Tyr Arg Asp
20 25 30

Thr Ala Ile Phe Leu Ala Gln Met Ser Leu Leu Glu Ser Thr Val Val
35 40 45

Ile Leu Leu Val Arg Leu Arg Thr
50 55

<210> 287
<211> 77
<212> PRT
<213> Homo sapien

<400> 287

Met Leu Leu Ala Val Arg Thr Thr Val Ile Cys Leu Gln Ser Cys Cys
1 5 10 15

Cys Arg Ile Gln Arg Thr Ala Thr Ile Thr Leu Asn Cys Phe Ala Leu
20 25 30

Ser Ser Ile Phe Asp Tyr Tyr Ile Ser His Asn Ile Thr Ile Ser His
35 40 45

Ser Ser Asn Tyr Ser Ala Gln Ile His Glu His Val Pro Ala Arg Ala
50 55 60

Ala Ala Arg Ser Ile Thr Trp Arg Arg Ser Ala Cys Ile
65 70 75

<210> 288
<211> 45
<212> PRT
<213> Homo sapien

<400> 288

Met Tyr Leu Gly Gln Leu Gly Asn His Arg Leu Lys Lys Leu Thr Leu
1 5 10 15

Val Ile Thr Arg Val Val Ser Asp Tyr Lys Gln His Ile Ile Asn Pro
20 25 30

Thr Ala Leu Ile Leu Ala Gln Arg Gln Asn Trp Thr Phe
35 40 45

<210> 289
<211> 44
<212> PRT
<213> Homo sapien

<400> 289

Met Lys Ala Leu Leu Cys Phe Leu Phe Tyr Ser Asp His Gln Thr Asp
1 5 10 15

Leu Ala Thr Leu Ile Val Lys Asn Glu Pro His Ser Ser Pro Gly Leu
20 25 30

Gly Leu Trp Arg Glu Met Asn Phe Leu Leu Glu Met
35 40

<210> 290
<211> 50
<212> PRT
<213> Homo sapien

<400> 290

Met Phe Arg Thr Ser Ser Tyr Arg Leu Leu Ile Tyr Lys Val Pro Val
1 5 10 15

Ala Val Thr Pro Thr Arg Lys Thr Trp Asn Cys Lys Gln Ala Gly Val
20 25 30

Thr Ser Val Thr Ser Asp Thr Val Gln Pro Glu Val Arg Phe Leu Phe
35 40 45

Trp Gly
50

<210> 291
<211> 44
<212> PRT
<213> Homo sapien

<400> 291

Met Ser Gln Trp Pro Val Ala Ser Lys Leu Val Gly Lys Glu Lys Thr
1 5 10 15

Phe Leu Phe Lys Gln Arg Lys Gly Phe Gly Glu Lys Thr Gly Ser Gly
20 25 30

Ser Gly Glu Val Phe Val Met Leu Gly Asp Arg Leu
35 40

<210> 292
<211> 61
<212> PRT

<213> Homo sapien

<400> 292

Met Val His Tyr Arg Lys Glu Lys Lys Thr Ser Val Ser Glu Trp Gln
1 5 10 15

Ile Leu Ile Ile Cys Ser Ser His Leu Phe Ser Ser Glu Asn His Ile
20 25 30

Thr Pro Glu Tyr Leu Pro Gly Arg Ile His His Thr Ala Pro Leu Glu
35 40 45

Pro Ala Ser Lys Asp Pro Phe Ala His Ile Val Ile Leu
50 55 60

<210> 293
<211> 112
<212> PRT
<213> Homo sapien

<400> 293

Met Gly Ile Ile Leu Asn Trp Leu Asn Gln Trp Ala Gln Ile Thr Tyr
1 5 10 15

Leu Pro Ser Leu Leu Cys Asp Ser Pro Ala Val Thr His Thr Ile His
20 25 30

Ile Leu Cys Thr Ser Asn Glu Gln Thr Trp Phe Pro Cys Phe Leu Asp
35 40 45

Ile Ser Met Thr Val Ser His Thr Asn Tyr Trp Val Arg Phe Phe Ser
50 55 60

Cys Tyr Arg Pro Thr Ser Cys Cys Leu Cys Val Val Leu Gln Lys Leu
65 70 75 80

Ser Ile Pro Thr Pro Leu Leu Cys His Leu Gln Glu Ser Gly Ile Val
85 90 95

Arg Ser Gln Leu Arg Lys Val Leu Val Pro Leu Thr Gly His Ile Leu
100 105 110

<210> 294
<211> 55
<212> PRT
<213> Homo sapien

<400> 294

Met Arg Phe Ile Phe Ile Cys Lys Pro Arg Gly Leu Ile Ile Leu Ile
1 5 10 15

Leu Tyr Glu Tyr Thr Cys Val Leu Gly Lys Ala Phe Ile Gln Gln Met
20 25 30

Pro Thr Thr Tyr Ser Val Pro Arg Pro Arg His Pro Val Thr Ser Trp
35 40 45

Arg Pro Ala Arg Ala Cys Ile
50 55

<210> 295
<211> 77
<212> PRT
<213> Homo sapien

<400> 295

Met Leu Glu Leu Pro Thr Phe Ser Phe Phe Phe Phe Gly Asp Arg Ala
1 5 10 15

Ser Leu Cys His Pro Gly Trp Ser Ala Gly Ala Ser Ser Leu Thr His
20 25 30

Leu Gln Pro Ser Phe Leu Pro Trp Gly Ala Gly Arg Phe Ser Cys Ala
35 40 45

Leu Gln Pro Pro Ser Leu Ala Gly Ile Tyr Arg Ala Leu Leu Gln Val
50 55 60

Ser His Ile Phe Ser Glu Lys Phe Leu Asn Trp Pro Pro
65 70 75

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/04197

A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) :C07H 21/02, 21/04
US CL :536/23.1, 24.3
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 24.3

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Search the USPTOs database of nucleic acid and protein sequences.

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	VINCENT et al. Oligonucleotides as short as 7-mers can be used for PCR amplification. DNA and Cell Biology. 1994, Vol. 13, No. 1, pages 75-82, note the octomer in the legend of Figure 5 and the heptamer in the legend of Figure 5 and compare to positions 60-67 of SEQ ID NO: 1.	1-5 and 7-8
X	SOMMER et al. Minimal homology requirements for PCR primers. Nucleic Acids Research. 1989, Vol. 17, No. 16, page 6749, note the first primer listed in Table I and compare the 3 nucleotides on its 3' end to the complementary nucleotides at positions 252-254 of SEQ ID NO: 1.	1-5 and 7-8

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

*	Special categories of cited documents:	"T"	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"A"	document defining the general state of the art which is not considered to be of particular relevance		
"E"	earlier document published on or after the international filing date	"X"	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"L"	document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y"	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"O"	document referring to an oral disclosure, use, exhibition or other means		
"P"	document published prior to the international filing date but later than the priority date claimed	"&"	document member of the same patent family

Date of the actual completion of the international search 31 MAY 2002	Date of mailing of the international search report 28 JUN 2002
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer ETHAN WHISENANT, PH.D. Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet) (July 1998)★

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US02/04197

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-5 and 7-8, SEQ ID NO: 1

Remark on Protest ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING
This ISA found multiple inventions as follows:

1. This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

I. **Claim(s) 1-5, 7-8,** drawn to polynucleotides, vectors comprising the polynucleotide, methods of introducing the polynucleotide to host cells, and host cells comprising the polynucleotide. This group comprises at least 171 different embodiments. See the requirement to elect a single sequence below.

II. **Claim(s) 10-11,** drawn to an isolated polypeptide.This group comprises at least 171 different embodiments. See the requirement to elect a single sequence below.

III. **Claim(s) 12,** drawn to antibodies. This group comprises at least 171 different embodiments. See the requirement to elect a single sequence below.

IV. **Claim(s) 9,** drawn to a method of synthesizing a polypeptide, classified in class 435, subclass 69.1. This group comprises at least 171 different embodiments. See the requirement to elect a single sequence below.

V. **Claim(s) 6 and 14,** drawn to a diagnostic method of using the polynucleotide of Claim 1. This group comprises at least 171 different embodiments. See the requirement to elect a single sequence below.

VI. **Claim(s) 13-14,** drawn to a diagnostic method of using a polypeptide. This group comprises at least 171 different embodiments. See the requirement to elect a single sequence below.

VII. **Claim(s) 15,** drawn to a kit for determining the presence of the nucleic acid molecule of claim 1 or the polypeptide of Claim 11. This group comprises at least 171 different embodiments. See the requirement to elect a single sequence below.

VIII. **Claim(s) 16,** drawn to a therapeutic method of using a polypeptide. This group comprises at least 171 different embodiments. See the requirement to elect a single sequence below.

IX. **Claim(s) 17,** drawn to a vaccine comprising the polypeptide of Claim 11 or a polynucleotide encoding said polypeptide of Claim 11. This group comprises at least 171 different embodiments. See the requirement to elect a single sequence below.

2. **Sequence Election Requirement Applicable to All Groups**
In addition, each Group detailed above reads on distinct sequences. Each sequence is distinct because they are unrelated sequences, and a further restriction is applied to each Group. For an selected Group drawn to amino acid sequences, the Applicants must elect a **single** amino acid sequence. For an elected Group drawn to nucleotide sequences, the Applicants must elect **one** nucleic acid sequence.
Examination will be restricted to only the elected sequence.

3. The inventions listed as Groups I-IX do not relate to a single

general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical feature(s).

The claims as drawn are related to each other because of the product i.e. the isolated nucleic acid molecule of Claim 1. However, since the isolated nucleic acid molecule of Claim 1, as claimed, is known, the claims are no longer linked by a special technical feature, because, by definition, the special technical feature must distinguish over the prior art. Without the special technical feature the claims lack unity.